GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: February 16, 2002, 14:26:07 ; Search time 1450.75 Seconds
(without alignments)
2945.214 Million cell updates/sec

Title: US-09-587-653-1
Perfect score: 259
Sequence: 1 aguuuggcgaccauggugga.....cccgcuuggaauuaaaaacu 259

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%

Result No.		Score	Query Match	Length	DB	ID	Description
1		259	100.0	9399	6	AX057392	AX057392 Sequence
2		259	100.0	9399	14	AF179612	AF179612 Hepatitis
3		257.4	99.4	259	6	AX055778	AX055778 Sequence
4	c	257.4	99.4	259	6	AX055779	AX055779 Sequence
5		257.4	99.4	259	6	AX055780	AX055780 Sequence
6	c	257.4	99.4	259	6	AX055781	AX055781 Sequence
7		257.4	99.4	309	6	AX055777	AX055777 Sequence
8		257.4	99.4	357	14	HGB18973	Y18973 Hepatitis G
9		257.4	99.4	9397	14	HGB277947	AJ277947 Hepatitis
10		82	31.7	82	6	AX055790	AX055790 Sequence
11		33.2	12.8	185516	9	AC004485	AC004485 Homo sapi
12		33.2	12.8	289563	2	AL450993	AL450993 Homo sapi
13	c	32.6	12.6	172751	10	CNS07EGP	AL591826 BAC 13C18
14		32.4	12.5	2354	14	PPM76VP1	D00056 Monkey B-ly
15		32.4	12.5	2354	14	PPMVP1M	M14494 Monkey B-ly
16		32.2	12.4	185273	2	AC019038	AC019038 Homo sapi
17	c	32	12.4	3179	6	AX179748	AX179748 Sequence
18	c	31.8	12.3	185302	2	AC010177	AC010177 Homo sapi
19		31.8	12.3	191877	9	AC007207	AC007207 Homo sapi
20		31.8	12.3	210004	9	AC069240	AC069240 Homo sapi
21		31.6	12.2	4498	3	AB030450	AB030450 Drosophil
22		31.4	12.1	1257	4	OCIGHENH	X13700 Rabbit IgH
23		31.4	12.1	137023	2	AC091091	AC091091 Gallus ga
24		31	12.0	177990	2	AC016385	AC016385 Homo sapi
25	c	31	12.0	205803	2	AC015802	AC015802 Homo sapi
26		31	12.0	206400	10	AC087184	AC087184 Mus Muscu
27		30.8	11.9	2354	14	PPMVP1	L29432 Monkey B-ly
28		30.8	11.9	2354	14	PPMVP1A	D00055 Monkey B-ly
29	c	30.8	11.9	5089	14	PPMCG	M30540 Monkey B-ly
30		30.8	11.9	5270	14	PPL	K02562 Lymphotropi
31	c	30.8	11.9	84904	9	AP001051	AP001051 Homo sapi
32	c	30.8	11.9	340000	9	AP001751	AP001751 Homo sapi
33		30.6	11.8	37762	3	AC005836	AC005836 Leishmani
34	c	30.6	11.8	159030	2	AC091729	AC091729 Homo sapi
35		30.6	11.8	172438	2	AC015711	AC015711 Homo sapi
36	c	30.6	11.8	180359	2	AC022064	AC022064 Homo sapi
37		30.6	11.8	185255	9	AC068181	AC068181 Homo sapi
38		30.6	11.8	219218	2	AC020817	AC020817 Mus muscu
39		30.4	11.7	21420	1	SCF85	AL110470 Streptomy
40		30.4	11.7	154279	2	AL138781	AL138781 Homo sapi
41		30.4	11.7	189121	2	AC022285	AC022285 Homo sapi
42		30.4	11.7	193335	2	AC016826	AC016826 Homo sapi
43	c	30.2	11.7	71578	2	AC025333	AC025333 Homo sapi
44		30.2	11.7	158519	9	AL135926	AL135926 Human DNA
45		30.2	11.7	162264	2	AC023787	AC023787 Homo sapi

ALIGNMENTS

RESULT 1
AX057392
LOCUS AX057392 9399 bp DNA PAT 17-JAN-2001
DEFINITION Sequence 1 from Patent WO0075337.
ACCESSION AX057392
VERSION AX057392.1 GI:12310132
KEYWORDS .
SOURCE Hepatitis GB virus B.
ORGANISM Hepatitis GB virus B
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae.
REFERENCE 1 (bases 1 to 9399)
AUTHORS Bukh,J., Yanagi,M., Emerson,S.U. and Purcell,R.H.
TITLE Infectious cdna clone of gb virus b and uses thereof
JOURNAL Patent: WO 0075337-A 1 14-DEC-2000;
THE GOVERNMENT OF THE UNITED STATES OF AMERICA (US)
FEATURES Location/Qualifiers
source 1. .9399
/organism="Hepatitis GB virus B"
/db_xref="taxon:39113"
BASE COUNT 2125 a 2326 c 2430 g 2518 t

```
ORIGIN

  Query Match            100.0%;  Score 259;  DB 6;  Length 9399;
  Best Local Similarity   76.1%;  Pred. No. 3.4e-67;
  Matches  197;  Conservative   62;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
          ||:::|||||||||:||:|||:|||||||||:::||||:||||||||:||:|:|||||||||:|
Db   9141 AGTTTGGCGACCATGGTGGATCAGAACCGTTTCGGGTGAAGCCATGGTCTGAAGGGGATG 9200

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          |||:|||::|:|||:||:|||||||||||||:|:||||:|||:|||||||:||:|||:|:|:
Db   9201 ACGTCCCTTCTGGCTCATCCACAAAAACCGTCTCGGGTGGGTGAGGAGTCCTGGCTGTGT 9260

Qy    121 gggaagcagucaguauaauucccgucgugugguggugacgccucacgacguauuuguccgc 180
          |||||||||:|||:|:||::||||:||:|:|:||:|||||||:|||||||:|:::|:||||
Db   9261 GGGAAGCAGTCAGTATAATTCCCGTCGTGTGTGGTGACGCCTCACGACGTATTTGTCCGC 9320

Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          :|:|||||||:||:|||||||||:|||||||||:::::|::|||||||||||||||||||
Db   9321 TGTGCAGAGCGTAGTACCAAGGGCTGCACCCCGGTTTTTGTTCCAAGCGGAGGGCAACCC 9380

Qy    241 ccgcuuggaauuaaaaacu 259
          ||||::|||||::||||||:
Db   9381 CCGCTTGGAATTAAAAACT 9399


RESULT   2
AF179612
LOCUS       AF179612      9399 bp    RNA             VRL       07-OCT-1999
DEFINITION  Hepatitis GB virus B polyprotein gene, complete cds.
ACCESSION   AF179612
VERSION     AF179612.1  GI:6014504
KEYWORDS    .
SOURCE      Hepatitis GB virus B.
  ORGANISM  Hepatitis GB virus B
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae.
REFERENCE   1  (bases 1 to 9399)
  AUTHORS   Bukh,J., Apgar,C.L. and Yanagi,M.
  TITLE     Toward a surrogate model for hepatitis C virus: An infectious
            molecular clone of the GB virus-B hepatitis agent
  JOURNAL   Virology 262 (2), 470-478 (1999)
  MEDLINE   99434247
REFERENCE   2  (bases 1 to 9399)
  AUTHORS   Bukh,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-1999) Hepatitis Viruses Section, Laboratory of
            Infectious Diseases, National Institute of Allergy and Infectious
            Diseases, National Institutes of Health, Building 7, Room 201, 7
            Center Dr. 0740, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .9399
                     /organism="Hepatitis GB virus B"
                     /specific_host="unknown"
                     /db_xref="taxon:39113"
                     /note="most closely related to hepatitis C virus;
                     experimental infection of tamarins causes acute resolving
                     hepatitis"
     CDS             446. .9040
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAF01368.1"
                     /db_xref="GI:6014505"
                     /translation="MPVISTQTSPVPAPRTRKNKQTQASYPVSIKTSVERGQRAKRKV
                     QRDARPRNYKIAGIHDGLQTLAQAALPAHGWGRQDPRHKSRNLGILLDYPLGWIGDVT
                     THTPLVGPLVAGAVVRPVCQIVRLLEDGVNWATGWFGVHLFVVCLLSLACPCSGARVT
                     DPDTNTTILTNCCQRNQVIYCSPSTCLHEPGCVICADECWVPANPYISHPSNWTGTDS
                     FLADHIDFVMGALVTCDALDIGELCGACVLVGDWLVRHWLIHIDLNETGTCYLEVPTG
                     IDPGFLGFIGWMAGKVEAVIFLTKLASQVPYAIATMFSSVHYLAVGALIYYASRGKWY
                     QLLLALMLYIEATSGNPIRVPTGCSIAEFCSPLMIPCPCHSYLSENVSEVICYSPKWT
                     RPITLEYNNSISWYPYTIPGARGCMVKFKNNTWGCCRIRNVPSYCTMGTDAVWNDTRN
                     TYEACGVTPWLTTAWHNGSALKLAILQYPGSKEMFKPHNWMSGHLYFEGSDTPIVYFY
                     DPVNSTLLPPERWARLPGTPPVVRGSWLQVPQGFYSDVKDLATGLITKDKAWKNYQVL
                     YSATGALSLTGVTTKAVVLILLGLCGSKYLILAYLCYLSLCFGRASGYPLRPVLPSQS
                     YLQAGWDVLSKAQVAPFALIFFICCYLRCRLRYAALLGFVPMAAGLPLTFFVAAAAAQ
                     PDYDWWVRLLVAGLVLWAGRNRGHRIALLVGPWPLVALLTLLHLVTPASAFDTEIIGG
                     LTIPPVVALVVMSRFGFFAHLLPRCALVNSYLWQRWENWFWNVTLRPERFFLVLVCFP
                     GATYDALVTFCVCHVALLCLTSSAASFFGTDSRVRAHRMLVRLGKCHAWYSHYVLKFF
                     LLVFGENGVFFYKHLHGDVLPNDFASKLPLQEPFFPFEGKARVVRNEGRRLACGDTVD
                     GLPVVARLGDLVFAGLAMPPDGWAITAPFTLQCLSERGTLSAMAVVMTGIDPRTWTGT
                     IFRLGSLATSYMGFVCDNVLYTAHHGSKGRRLAHPTGSIHPITVDAANDQDIYQPPCG
                     AGSLTRCSCGETKGYLVTRLGSLVEVNKSDDPYWCVCGALPMAVAKGSSGAPILCSSG
                     HVIGMFTAARNSGGSVSQIRVRPLVCAGYHPQYTAHATLDTKPTVPNEYSVQILIAPT
                     GSGKSTKLPLSYMQEKYEVLVLNPSVATTASMPKYMHATYGVNPNCYFNGKCTNTGAS
                     LTYSTYGMYLTGACSRNYDVIICDECHATDATTVLGIGKVLTEAPSKNVRLVVLATAT
                     PPGVIPTPHANITEIQLTDEGTIPFHGKKIKEENLKKGRHLIFEATKKHCDELANELA
                     RKGITAVSYYRGCDISKIPEGDCVVVATDALCTGYTGDFDSVYDCSLMVEGTCHVDLD
                     PTFTMGVRVCGVSAIVKGQRRGRTGRGRAGIYYYVDGSCTPSGMVPECNIVEAFDAAK
                     AWYGLSSTEAQTILDTYRTQPGLPAIGANLDEWADLFSMVNPEPSFVNTAKRTADNYV
                     LLTAAQLQLCHQYGYAAPNDAPRWQGARLGKKPCGVLWRLDGADACPGPEPSEVTRYQ
                     MCFTEVNTSGTAALAVGVGVAMAYLAIDTFGATCVRRCWSITSVPTGATVAPVVDEEE
                     IVEECASFIPLEAMVAAIDKLKSTITTTSPFTLETALAEKLNTFLGPHAATILAIIEYC
                     CGLVTLPDNPFASCVFAFIAGITTPLPHKIKMFLSLFGGAIASKLTDARGALAFMMAG
                     AAGTALGTWTSVGFVFDMLGGYAAASSTACLTFKCLMGEWPTMDQLAGLVYSAFNPAA
                     GVVGVLSACAMFALTTAGPDHWPNRLLTMLARSNTVCNEYFIATRDIRRKILGILEAS
                     TPWSVISACIRWLHTPTEDDCGLIAWGLEIWQYVCNFFVICFNVLKAGVQSMVNIPGC
                     PFYSCQKGYKGPWIGSGMLQARCPCGAELIFSVENGFAKLYKGPRTCSNYWRGAVPVN
                     ARLCGSARPDPTDWTSLVVNYGVRDYCKYEKMGDHIFVTAVSSPNVCFTQVPPTLRAA
                     VAVDGVQVQCYLGEPKTPWTTSACCYGPDGKGKTVKLPFRVDGHTPGVRMQLNLRDAL
                     ETNDCNSTNNTPSDEAAVSALVFKQELRRTNQLLEAISAGVDTTKLPAPSIEEVVVRK
                     RQFRARTGSLTLPPPPRSVPGVSCPESLQRSDPLEGPSNLPPSPPVLQLAMPMPLLGA
                     GECNPFTAIGCAMTETGGGPDDLPSYPPKKEVSEWSDESWSTATTASSYVTGPPYPKI
                     RGKDSTQSAPAKRPTKKKLGKSEFSCSMSYTWTDVISFKTASKVLSATRAITSGFLKQ
                     RSLVYVTEPRDAELRKQKVTINRQPLFPPSYHKQVRLAKEKASKVVGVMWDYDEVAAH
                     TPSKSAKSHITGLRGTDVRSGAARKAVLDLQKCVEAGEIPSHYRQTVIVPKEEVFVKT
                     PQKPTKKPPRLISYPHLEMRCVEKMYYGQVAPDVVKAVMGDAYGFVDPRTRVKRLLSM
                     WSPDAVGATCDTVCFDSTITPEDIMVETDIYSAAKLSDQHRAGIHTIARQLYAGGPMI
                     AYDGREIGYRRCRSSGVYTTSSSNSLTCWLKVNAAAEQAGMKNPRFLICGDDCTVIWK
                     SAGADADKQAMRVFASWMKVMGAPQDCVPQPKYSLEELTSCSSNVTSGITKSGKPYYF
                     LTRDPRIPLGRCSAEGLGYNPSAAWIGYLIHHYPCLWVSRVLAVHFMEQMLFEDKLPE
                     TVTFDWYGKNYTVPVEDLPSIIAGVHGIEAFSVVRYTNAEILRVSQSLTDMTMPPLRA
                     WRKKARAVLASAKRRGGAHAKLARFLLWHATSRPLPDLDKTSVARYTTFNYCDVYSPE
                     GDVFITPQRRLQKFLVKYLAVIVFALGLIAVGLAIS"
BASE COUNT     2125 a   2326 c   2430 g   2518 t
ORIGIN

  Query Match            100.0%;  Score 259;  DB 14;  Length 9399;
  Best Local Similarity   76.1%;  Pred. No. 3.4e-67;
  Matches  197;  Conservative   62;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
          ||:::|||||||||:||:|||:|||||||||:::||||:||||||||:||:|:|||||||||:|
Db   9141 AGTTTGGCGACCATGGTGGATCAGAACCGTTTCGGGTGAAGCCATGGTCTGAAGGGGATG 9200

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          |||:|||::|:|||:||:|||||||||||||:|:||||:|||:|||||||:||:|||:|:|:
Db   9201 ACGTCCCTTCTGGCTCATCCACAAAAACCGTCTCGGGTGGGTGAGGAGTCCTGGCTGTGT 9260

Qy    121 gggaagcagucaguauaauucccgucgugugguggugacgccucacgacguauuuguccgc 180
          |||||||||:|||:|:||::||||:||:|:|:||:|||||||:|||||||:|:::|:||||
Db   9261 GGGAAGCAGTCAGTATAATTCCCGTCGTGTGTGGTGACGCCTCACGACGTATTTGTCCGC 9320

Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          :|:|||||||||:||:|||||||||:|||||||||:::::|::|||||||||||||||||||
Db   9321 TGTGCAGAGCGTAGTACCAAGGGCTGCACCCCGGTTTTTGTTCCAAGCGGAGGGCAACCC 9380

Qy    241 ccgcuuggaauuaaaaacu 259
          ||||::|||||::||||||:
Db   9381 CCGCTTGGAATTAAAAACT 9399


RESULT   3
AX055778
LOCUS       AX055778      259 bp    mRNA            PAT       13-JAN-2001
```

```
DEFINITION  Sequence 2 from Patent WO0073466.
ACCESSION   AX055778
VERSION     AX055778.1  GI:12228890
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 259)
  AUTHORS   Traboni,C.I.
  TITLE     Novel gbv sequence
  JOURNAL   Patent: WO 0073466-A 2 07-DEC-2000;
            ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI; S.P.A.
            (IT)
FEATURES             Location/Qualifiers
     source          1. .259
                     /organism="unidentified"
                     /db_xref="taxon:32644"
                     /note="GBV-B-like virus"
BASE COUNT       53 a     64 c     81 g     61 t
ORIGIN
```

Query Match 99.4%; Score 257.4; DB 6; Length 259;
Best Local Similarity 76.1%; Pred. No. 2e-66;
Matches 197; Conservative 61; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
          ||:::|||||||||:||:|||:||||||||:::||||:|||||||:||:|:|||||||:|
Db      1 AGTTTGGCGACCATGGTGGATCAGAACCGTTTCGGGTGAAGCCATGGTCTGAAGGGGATG 60

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          |||:|||::|:|||:||:|||||||||||||:|:||||:|||:||||||:||:|||:|:|:
Db     61 ACGTCCCTTCTGGCTCATCCACAAAAACCGTCTCGGGTGGGTGAGGAGTCCTGGCTGTGT 120

Qy    121 gggaagcagucaguauaauuucccgucgugugguggugacgccucacgacguauuuguccgc 180
          |||||||||:|||:|:||::||||:||:|:|:||:|||||:|||||||:| ::|:||||
Db    121 GGGAAGCAGTCAGTATAATTCCCGTCGTGTGTGGTGACGCCTCACGACGTACTTGTCCGC 180

Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          :|:|||||||:||:|||||||||:|||||||||:::::|::|||||||||||||||||||
Db    181 TGTGCAGAGCGTAGTACCAAGGGCTGCACCCCGGTTTTTGTTCCAAGCGGAGGGCAACCC 240

Qy    241 ccgcuuggaauuaaaaacu 259
          ||||::||||::||||||:
Db    241 CCGCTTGGAATTAAAAACT 259
```

```
RESULT   4
AX055779/c
LOCUS       AX055779      259 bp    mRNA            PAT       13-JAN-2001
DEFINITION  Sequence 3 from Patent WO0073466.
ACCESSION   AX055779
VERSION     AX055779.1  GI:12228891
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 259)
  AUTHORS   Traboni,C.I.
  TITLE     Novel gbv sequence
  JOURNAL   Patent: WO 0073466-A 3 07-DEC-2000;
            ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI; S.P.A.
            (IT)
FEATURES             Location/Qualifiers
     source          1. .259
                     /organism="unidentified"
                     /db_xref="taxon:32644"
                     /note="GBV-B-like virus"
BASE COUNT       61 a     81 c     64 g     53 t
ORIGIN
```

Query Match 99.4%; Score 257.4; DB 6; Length 259;
Best Local Similarity 76.1%; Pred. No. 2e-66;
Matches 197; Conservative 61; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
          ||:::|||||||||:||:|||:||||||||:::||||:|||||||:||:|:|||||||:|
Db    259 AGTTTGGCGACCATGGTGGATCAGAACCGTTTCGGGTGAAGCCATGGTCTGAAGGGGATG 200

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          |||:|||::|:|||:||:|||||||||||||:|:||||:|||:||||||:||:|||:|:|:
Db    199 ACGTCCCTTCTGGCTCATCCACAAAAACCGTCTCGGGTGGGTGAGGAGTCCTGGCTGTGT 140

Qy    121 gggaagcagucaguauaauuucccgucgugugguggugacgccucacgacguauuuguccgc 180
          |||||||||:|||:|:||::||||:||:|:|:||:|||||:|||||||:| ::|:||||
Db    139 GGGAAGCAGTCAGTATAATTCCCGTCGTGTGTGGTGACGCCTCACGACGTACTTGTCCGC 80

Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          :|:|||||||:||:|||||||||:|||||||||:::::|::|||||||||||||||||||
Db     79 TGTGCAGAGCGTAGTACCAAGGGCTGCACCCCGGTTTTTGTTCCAAGCGGAGGGCAACCC 20

Qy    241 ccgcuuggaauuaaaaacu 259
          ||||::||||::||||||:
Db     19 CCGCTTGGAATTAAAAACT 1
```

```
RESULT   5
AX055780
LOCUS       AX055780      259 bp    DNA             PAT       13-JAN-2001
DEFINITION  Sequence 4 from Patent WO0073466.
ACCESSION   AX055780
VERSION     AX055780.1  GI:12228892
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 259)
  AUTHORS   Traboni,C.I.
  TITLE     Novel gbv sequence
  JOURNAL   Patent: WO 0073466-A 4 07-DEC-2000;
            ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI; S.P.A.
            (IT)
FEATURES             Location/Qualifiers
     source          1. .259
                     /organism="unidentified"
                     /db_xref="taxon:32644"
                     /note="GBV-B-like virus"
     3'UTR           1. .259
                     /note="cDNA of positive strand 3'UTR of GBV-B"
BASE COUNT       53 a     64 c     81 g     61 t
ORIGIN
```

Query Match 99.4%; Score 257.4; DB 6; Length 259;
Best Local Similarity 76.1%; Pred. No. 2e-66;
Matches 197; Conservative 61; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
          ||:::|||||||||:||:|||:||||||||:::||||:|||||||:||:|:|||||||:|
Db      1 AGTTTGGCGACCATGGTGGATCAGAACCGTTTCGGGTGAAGCCATGGTCTGAAGGGGATG 60

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          |||:|||::|:|||:||:|||||||||||||:|:||||:|||:||||||:||:|||:|:|:
Db     61 ACGTCCCTTCTGGCTCATCCACAAAAACCGTCTCGGGTGGGTGAGGAGTCCTGGCTGTGT 120

Qy    121 gggaagcagucaguauaauuucccgucgugugguggugacgccucacgacguauuuguccgc 180
          |||||||||:|||:|:||::||||:||:|:|:||:|||||:|||||||:| ::|:||||
Db    121 GGGAAGCAGTCAGTATAATTCCCGTCGTGTGTGGTGACGCCTCACGACGTACTTGTCCGC 180

Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          :|:|||||||:||:|||||||||:|||||||||:::::|::|||||||||||||||||||
Db    181 TGTGCAGAGCGTAGTACCAAGGGCTGCACCCCGGTTTTTGTTCCAAGCGGAGGGCAACCC 240

Qy    241 ccgcuuggaauuaaaaacu 259
          ||||::||||::||||||:
```

```
Db    241 CCGCTTGGAATTAAAAACT 259


RESULT   6
AX055781/c
LOCUS       AX055781      259 bp    DNA             PAT       13-JAN-2001
DEFINITION  Sequence 5 from Patent WO0073466.
ACCESSION   AX055781
VERSION     AX055781.1  GI:12228893
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 259)
  AUTHORS   Traboni,C.I.
  TITLE     Novel gbv sequence
  JOURNAL   Patent: WO 0073466-A 5 07-DEC-2000;
            ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI; S.P.A.
            (IT)
FEATURES             Location/Qualifiers
     source          1. .259
                     /organism="unidentified"
                     /db_xref="taxon:32644"
                     /note="GBV-B-like virus"
     3'UTR           1. .259
                     /note="cDNA of negative strand 3'UTR of GBV-B"
BASE COUNT       61 a     81 c     64 g     53 t
ORIGIN


  Query Match             99.4%;  Score 257.4;  DB 6;  Length 259;
  Best Local Similarity   76.1%;  Pred. No. 2e-66;
  Matches  197;  Conservative   61;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
          ||:::|||||||||:||:|||:||||||||:::||||:|||||||:||:|:|||||||:|
Db    259 AGTTTGGCGACCATGGTGGATCAGAACCGTTTCGGGTGAAGCCATGGTCTGAAGGGGATG 200

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          |||:|||::|:|||:||:||||||||||||:|:||||:|||:||||||:||:|||:|:|:
Db    199 ACGTCCCTTCTGGCTCATCCACAAAAACCGTCTCGGGTGGGTGAGGAGTCCTGGCTGTGT 140

Qy    121 gggaagcagucaguauaauuucccgucgugugguggugacgccucacgacguauuuguccgc 180
          |||||||||:|||:|:||::||||:||:|:|:||:||||||:|||||||:| ::|:||||
Db    139 GGGAAGCAGTCAGTATAATTCCCGTCGTGTGTGGTGACGCCTCACGACGTACTTGTCCGC 80

Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          :|:|||||||:||:||||||||:|||||||||:::::|::|||||||||||||||||||
Db     79 TGTGCAGAGCGTAGTACCAAGGGCTGCACCCCGGTTTTTGTTCCAAGCGGAGGGCAACCC 20

Qy    241 ccgcuuggaauuaaaaacu 259
          ||||::||||::||||||:
Db     19 CCGCTTGGAATTAAAAACT 1


RESULT   7
AX055777
LOCUS       AX055777      309 bp    DNA             PAT       13-JAN-2001
DEFINITION  Sequence 1 from Patent WO0073466.
ACCESSION   AX055777
VERSION     AX055777.1  GI:12228889
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 309)
  AUTHORS   Traboni,C.I.
  TITLE     Novel gbv sequence
  JOURNAL   Patent: WO 0073466-A 1 07-DEC-2000;
            ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI; S.P.A.
            (IT)
FEATURES             Location/Qualifiers
     source          1. .309
                     /organism="unidentified"
                     /db_xref="taxon:32644"
                     /note="GBV-B-like virus"
     3'UTR           1. .309
                     /note="cDNA complementary to the last 49 published 3'UTR
                     nucleotides of GBV-B plus a novel nucleotide (C) in
                     position 44 and 259 novel nucleotide sequence at the 3'
                     end."
BASE COUNT       64 a     80 c    100 g     65 t
ORIGIN


  Query Match             99.4%;  Score 257.4;  DB 6;  Length 309;
  Best Local Similarity   76.1%;  Pred. No. 1.9e-66;
  Matches  197;  Conservative   61;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
          ||:::|||||||||:||:|||:||||||||:::||||:|||||||:||:|:|||||||:|
Db     51 AGTTTGGCGACCATGGTGGATCAGAACCGTTTCGGGTGAAGCCATGGTCTGAAGGGGATG 110

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          |||:|||::|:|||:||:||||||||||||:|:||||:|||:||||||:||:|||:|:|:
Db    111 ACGTCCCTTCTGGCTCATCCACAAAAACCGTCTCGGGTGGGTGAGGAGTCCTGGCTGTGT 170

Qy    121 gggaagcagucaguauaauuucccgucgugugguggugacgccucacgacguauuuguccgc 180
          |||||||||:|||:|:||::||||:||:|:|:||:||||||:|||||||:| ::|:||||
Db    171 GGGAAGCAGTCAGTATAATTCCCGTCGTGTGTGGTGACGCCTCACGACGTACTTGTCCGC 230

Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          :|:|||||||:||:||||||||:|||||||||:::::|::|||||||||||||||||||
Db    231 TGTGCAGAGCGTAGTACCAAGGGCTGCACCCCGGTTTTTGTTCCAAGCGGAGGGCAACCC 290

Qy    241 ccgcuuggaauuaaaaacu 259
          ||||::||||::||||||:
Db    291 CCGCTTGGAATTAAAAACT 309


RESULT   8
HGB18973
LOCUS       HGB18973      357 bp    RNA             VRL       19-NOV-1999
DEFINITION  Hepatitis GB virus B partial 3'UTR region.
ACCESSION   Y18973
VERSION     Y18973.1  GI:6018427
KEYWORDS    3' UTR.
SOURCE      Hepatitis GB virus B.
  ORGANISM  Hepatitis GB virus B
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae.
REFERENCE   1  (bases 1 to 357)
  AUTHORS   Sbardellati,A., Scarselli,E., Tomei,L., Kekule,A.S. and Traboni,C.
  TITLE     Identification of a novel sequence at the 3' end of the GB virus B
            genome
  JOURNAL   J. Virol. 73 (12), 10546-10550 (1999)
  MEDLINE   20027287
REFERENCE   2  (bases 1 to 357)
  AUTHORS   Traboni,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1999) C. Traboni, IRBM P.Angeletti, Ist. di
            Ricerche di Biologia Molecolare, Via Pontina. km. 30.600, 00040
            Pomezia, Rome, ITALY
FEATURES             Location/Qualifiers
     source          1. .357
                     /organism="Hepatitis GB virus B"
                     /specific_host="tamarin"
                     /db_xref="taxon:39113"
     3'UTR           1. .357
                     /evidence=experimental
     misc_feature    31. .48
                     /note="poly-U"
                     /evidence=experimental
     misc_feature    99. .357
                     /note="3' Y region"
                     /evidence=experimental
BASE COUNT       77 a     88 c    101 g     91 t
```

```
ORIGIN

  Query Match            99.4%;  Score 257.4;  DB 14;  Length 357;
  Best Local Similarity  76.1%;  Pred. No. 1.9e-66;
  Matches  197;  Conservative   61;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
           ||:::|||||||:||:|||:||||||||:::||||:|||||||:||:|:||||||||:|
Db      99 AGTTTGGCGACCATGGTGGATCAGAACCGTTTCGGGTGAAGCCATGGTCTGAAGGGGATG 158

Qy      61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
           |||:|||::|:|||:||:|||||||||||||:|:||||:|||:|||||||:||:|||:|:|:
Db     159 ACGTCCCTTCTGGCTCATCCACAAAAACCGTCTCGGGTGGGTGAGGAGTCCTGGCTGTGT 218

Qy     121 gggaagcagucaguauaauuucccgucgugugguggugacgccucacgacguauuuguccgc 180
           |||||||||:|||:|:||::||||:||:|:|:||:||||||:|||||||:| ::|:||||
Db     219 GGGAAGCAGTCAGTATAATTCCCGTCGTGTGTGGTGACGCCTCACGACGTACTTGTCCGC 278

Qy     181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
           :|:|||||||:||:|||||||||:||||||||:::::|::|||||||||||||||||||
Db     279 TGTGCAGAGCGTAGTACCAAGGGCTGCACCCCGGTTTTTGTTCCAAGCGGAGGGCAACCC 338

Qy     241 ccgcuuggaauuaaaaacu 259
           ||||::|||||::||||||:
Db     339 CCGCTTGGAATTAAAAACT 357


RESULT  9
HGB277947
LOCUS       HGB277947    9397 bp    RNA             VRL       26-FEB-2001
DEFINITION  Hepatitis GB virus B genomic RNA for polyprotein.
ACCESSION   AJ277947
VERSION     AJ277947.1  GI:13162187
KEYWORDS    ORF1; polyprotein.
SOURCE      Hepatitis GB virus B.
  ORGANISM  Hepatitis GB virus B
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae.
REFERENCE   1  (bases 1 to 9397)
  AUTHORS   Scarselli,E., Sbardellati,A. and Traboni,C.
  TITLE     Identification of an infectious GBV-B cDNA
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9397)
  AUTHORS   Traboni,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-2000) Traboni C., Biochemistry, IRBM P.
            Angeletti, via Pontina km. 30.600, 00040 Pomezia (Roma), ITALY
FEATURES             Location/Qualifiers
     source          1. .9397
                     /organism="Hepatitis GB virus B"
                     /db_xref="taxon:39113"
                     /clone="ACY/GBV-B/FL-3"
     5'UTR           1. .445
     CDS             446. .9040
                     /note="ORF1"
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="CAC33083.1"
                     /db_xref="GI:13162188"
                     /translation="MPVISTQTSPVPAPRTRKNKQTQASYPVSIKTSVERGQRAKRKV
                     QRDARPRNYKIAGIHDGLQTLAQAALPAHGWGRQDPRHKSRNLGILLDYPLGWIGDVT
                     THTPLVGPLVAGAVVRPVCQIVRLLEDGVNWATGWFGVHLFVVCLLSLACPCSGARVT
                     DPDTNTTILTNCCQRNQVIYCSPSTCLHEPGCVICADECWVPANPYISHPSNWTGTDS
                     FLADHIDFVMGALVTCDALDIGELCGACVLVGDWLVRHWLIHIDLNETGTCYLEVPTG
                     IDPGFLGFIGWMAGKVEAVIFLTKLASQVPYAIATMFSSVHYLAVGALIYYASRGKWY
                     QLLLALMLYIEATSGNPIRVPTGCSIAEFCSPLMIPCPCHSYLSENVSEVICYSPKWT
                     RPVTLEYNNSISWYPYTIPGARGCMVKFKNNTWGCCRIRNVPSYCTMGTDAVWNDTRN
                     TYEACGVTPWLTTAWHNGSALKLAILQYPGSKEMFKPHNWMSGHLYFEGSDTPIVYFY
                     DPVNSTLLPPERWARLPGTPPVVRGSWLQVPQGFYSDVKDLATGLITKDKAWKNYQVL
                     YSATGALSLTGVTTKAVVLILLGLCGSKYLILAYLCYLSLCFGRASGYPLRPVLPSQS
                     YLQAGWDVLSKAQVAPFALIFFICCYLRCRLRYAALLGFVPMAAGLPLTFFVAAAAAQ
                     PDYDWWVRLLVAGLVLWAGRDRGHRIALLVGPWPLVALLTLLHLVTPASAFDTEIIGG
                     LTIPPVVALVVMSRFGFFAHLLPRCALVNSYLWQRWENWFWNVTLRPERFFLVLVCFP
                     GATYDALVTFCVCHVALLCLTSSAASFFGTDSRVRAHRMLVRLGKCHAWYSHYVLKFF
                     LLVFGENGVFFYKHLHGDVLPNDFASKLPLQEPFFPFEGKARVYRNEGRRLACGDTVD
                     GLPVVARLGDLVFAGLAMPPDGWAITAPFTLQCLSERGTLSAMAVVMTGIDPRTWTGT
                     IFRLGSLATSYMGFVCDNVLYTAHHGSKGRRLAHPTGSIHPITVDAANDQDIYQPPCG
                     AGSLTRCSCGETKGYLVTRLGSLVEVNKSDDPYWCVCGALPMAVAKGSSGAPILCSSG
                     HVIGMFTAARNSGGSVSQIRVRPLVCAGYHPQYTAHATLDTKPTVPNEYSVQILIAPT
                     GSGKSTKLPLSYMQEKYEVLVLNPSVATTASMPKYMHATYGVNPNCYFNGKCTNTGAS
                     LTYSTYGMYLTGACSRNYDVIICDECHATDATTVLGIGKVLTEAPSKNVRLVVLATAT
                     PPGVIPTPHANITEIQLTDEGTIPFHGKKIKEENLKKGRHLIFEATKKHCDELANELA
                     RKGITAVSYYRGCDISKIPEGDCVVVATDALCTGYTGDFDSVYDCSLMVEGTCHVDLD
                     PTFTMGVRVCGVSAIVKGQRRGRTGRGRAGIYYYVDGSCTPSGMVPECNIVEAFDAAK
                     AWYGLSSTEAQTILDTYRTQPGLPAIGANLDEWADLFSMVNPEPSFVNTAKRTADNYV
                     LLTAAQLQLCHQYGYAAPNDAPRWQGARLGKKPCGVLWRLDGADACPGPEPSEVTRYQ
                     MCFTEVNTSGTAALAVGVGVAMAYLAIDTFGATCVRRCWSITSVPTGATVAPVVDEEE
                     IVEECASFIPLEAMVAAIDKLKSTITTSPFTLETALEKLNTFLGPHAATILAIIEYC
                     CGLVTLPDNPFASCVFAFIAGITTPLPHKIKMFLSLFGGAIASKLTDARGALAFMMAG
                     AAGTALGTWTSVGFVFDMLGGYAAASSTACLTFKCLMGEWPTMDQLAGLVYSAFNPAA
                     GVVGVLSACAMFALTTAGPDHWPNRLLTMLARSNTVCNEYFIATRDIRRKILGILEAS
                     TPWSVISACIRWLHTPTEDDCGLIAWGLEIWQYVCNFFVICFNVLKAGVQSMVNIPGC
                     PFYSCQKGYKGPWIGSGMLQARCPCGAELIFSVENGFAKLYKGPRTCSNYWRGAVPVN
                     ARLCGSARPDPTDWTSLVVNYGVRDYCKYEKLGDHIFVTAVSSPNVCFTQVPPTLRAA
                     VAVDGVQVQCYLGEPKTPWTTSACCYGPDGKGKTVKLPFRVDGHTPGVRMQLNLRDAL
                     ETNDCNSTNNTPSDEAAVSALVFKQELRRTNQLLEAISAGVDTTKLPAPSIEEVVVRK
                     RQFRARTGSLTLPPPPRSVPGVSCPESLQRSDPLEGPSNLPPSPPVLQLAMPMPLLGA
                     GECNPFTAIGCAMTETGGGPDDLPSYPPKKEVSEWSDESWSTATTASSYVTGPPYPKI
                     RGKDSTQSAPAKRPTKKKLGKSEFSCSMSYTWTDVISFKTASKVLSATRAITSGFLKQ
                     RSLVYVTEPRDAELRKQKVTINRQPLFPPSYHKQVRLAKEKASKVVGVMWDYDEVAAH
                     TPSKSAKSHITGLRGTDVRSGAARKAVLDLQKCVEAGEIPSHYRQTVIVPKEEVFVKT
                     PQKPTKKPPRLISYPHLEMRCVEKMYYGQVAPDVVKAVMGDAYGFVDPRTRVKRLLSM
                     WSPDAVGATCDTVCFDSTITPEDIMVETDIYSAAKLSDQHRAGIHTIARQLYAGGPMI
                     AYDGREIGYRRCRSSGVYTTSSSNSLTCWLKVNAAAEQAGMKNPRFLICGDDCTVIWK
                     SAGADADKQAMRVFASWMKVMGAPQDCVPQPKYSLEELTSCSSNVTSGITKSGKPYYF
                     LTRDPRIPLGRCSAEGLGYNPSAAWIGYLIHHYPCLWVSRVLAVHFMEQMLFEDKLPE
                     TVTFDWYGKNYTVPVEDLPSIIAGVHGIEAFSVVRYTNAEILRVSQSLTDMTMPPLRA
                     WRKKARAVLASAKRRGGAHAKLARFLLWHATSRPLPDLDKTSVARYTTFNYCDVYSPE
                     GDVFVTPQRRLQKFLVKYLAVIVFALGLIAVGLAIS"
     3'UTR           9041. .9397
BASE COUNT     2125 a    2331 c    2431 g    2510 t
ORIGIN


  Query Match            99.4%;  Score 257.4;  DB 14;  Length 9397;
  Best Local Similarity  76.1%;  Pred. No. 1e-66;
  Matches  197;  Conservative   61;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
           ||:::|||||||:||:|||:||||||||:::||||:|||||||:||:|:||||||||:|
Db    9139 AGTTTGGCGACCATGGTGGATCAGAACCGTTTCGGGTGAAGCCATGGTCTGAAGGGGATG 9198

Qy      61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
           |||:|||::|:|||:||:|||||||||||||:|:||||:|||:|||||||:||:|||:|:|:
Db    9199 ACGTCCCTTCTGGCTCATCCACAAAAACCGTCTCGGGTGGGTGAGGAGTCCTGGCTGTGT 9258

Qy     121 gggaagcagucaguauaauuucccgucgugugguggugacgccucacgacguauuuguccgc 180
           |||||||||:|||:|:||::||||:||:|:|:||:||||||:|||||||:| ::|:||||
Db    9259 GGGAAGCAGTCAGTATAATTCCCGTCGTGTGTGGTGACGCCTCACGACGTACTTGTCCGC 9318

Qy     181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
           :|:|||||||:||:|||||||||:||||||||:::::|::|||||||||||||||||||
Db    9319 TGTGCAGAGCGTAGTACCAAGGGCTGCACCCCGGTTTTTGTTCCAAGCGGAGGGCAACCC 9378

Qy     241 ccgcuuggaauuaaaaacu 259
           ||||::|||||::||||||:
Db    9379 CCGCTTGGAATTAAAAACT 9397


RESULT  10
AX055790
LOCUS       AX055790       82 bp    mRNA            PAT       13-JAN-2001
DEFINITION  Sequence 14 from Patent WO0073466.
ACCESSION   AX055790
VERSION     AX055790.1  GI:12228902
KEYWORDS    .
```

```
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 82)
  AUTHORS   Traboni,C.I.
  TITLE     Novel gbv sequence
  JOURNAL   Patent: WO 0073466-A 14 07-DEC-2000;
            ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI; S.P.A.
            (IT)
FEATURES             Location/Qualifiers
     source          1. .82
                     /organism="unidentified"
                     /db_xref="taxon:32644"
                     /note="GBV-B-like virus"
BASE COUNT       19 a     23 c     23 g     17 t
ORIGIN


  Query Match             31.7%;  Score 82;  DB 6;  Length 82;
  Best Local Similarity   79.3%;  Pred. No. 8.9e-14;
  Matches   65;  Conservative   17;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    178 cgcugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaa 237
          |||:|:||||||||:||:||||||||||:||||||||||:::::|::|||||||||||||||||
Db      1 CGCTGTGCAGAGCGTAGTACCAAGGGCTGCACCCCGGTTTTTGTTCCAAGCGGAGGGCAA 60

Qy    238 cccccgcuuggaauuaaaaacu 259
          |||||||::||||::|||||||:
Db     61 CCCCCGCTTGGAATTAAAAACT 82


RESULT  11
AC004485
LOCUS       AC004485   185516 bp    DNA             PRI       03-FEB-2000
DEFINITION  Homo sapiens BAC clone CTB-118E13 from 7p15-p21, complete sequence.
ACCESSION   AC004485
VERSION     AC004485.1  GI:2992497
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 185516)
  AUTHORS   Lacy,M.
  TITLE     The sequence of Homo sapiens BAC clone CTB-118E13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 185516)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   3  (bases 1 to 185516)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     -------------- Genome Center
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu/gsc
              Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
              Center project name: H_RG118E13
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            Clone CTB-118E13 is from the first release of the human BAC library
            CITB-978SK-B.  The library contains cloned DNA from the male
            fibroblast cell line 978SK.  See:  Shizuya et al., Proc. Natl.
            Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8
            (1996).  This clone is available from Research Genetics, Inc.
            (http://www.resgen.com).
            VECTOR:  pBeloBAC11
            Selection:  chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            The actual start of this clone is at base position 1 of CTB-118E13;
            actual end is at 185516 of CTB-118E13. The orientation of this
            clone is unknown.

            This clone contains STS sWSS1475 (NID:g1222787).
FEATURES             Location/Qualifiers
     source          1. .185516
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7p15-p21"
                     /clone="CTB-118E13"
                     /clone_lib="CITB-978SK-B"
     repeat_region   1376. .2438
                     /rpt_family="L1"
     repeat_region   3388. .3454
                     /rpt_family="MER1_type"
     repeat_region   4144. .4443
                     /rpt_family="Alu"
     repeat_region   4454. .4792
                     /rpt_family="L2"
     repeat_region   5182. .5483
                     /rpt_family="Alu"
     repeat_region   6636. .6800
                     /rpt_family="MER1_type"
     repeat_region   7781. .7854
                     /rpt_family="Retroviral"
     repeat_region   7924. .8143
                     /rpt_family="Retroviral"
     repeat_region   9235. .9285
                     /rpt_family="MIR"
     repeat_region   10636. .10813
                     /rpt_family="L1"
     repeat_region   11664. .11841
                     /rpt_family="MIR"
     repeat_region   13662. .13846
                     /rpt_family="L1"
     repeat_region   14191. .14355
                     /rpt_family="MIR"
     repeat_region   14838. .14948
                     /rpt_family="MER81"
     repeat_region   15640. .15820
                     /rpt_family="L2"
     repeat_region   17476. .17550
                     /rpt_family="Mariner"
     repeat_region   17590. .17892
                     /rpt_family="Alu"
     repeat_region   18483. .18564
```

```
                     /rpt_family="L2"
     repeat_region   18690. .18980
                     /rpt_family="Alu"
     repeat_region   19570. .19699
                     /rpt_family="L2"
     repeat_region   20714. .20817
                     /rpt_family="MIR"
     repeat_region   21043. .21384
                     /rpt_family="MER2_type"
     repeat_region   22077. .22266
                     /rpt_family="MER1_type"
     misc_feature    22605. .22805
                     /note="match to EST AA502194 (NID:g2237161) ng73e12.s1"
     misc_feature    complement(22605. .22805)
                     /note="match to EST AA226171 (NID:g1847487) nc11g07.r1"
     misc_feature    complement(22609. .22808)
                     /note="match to EST AA484674 (NID:g2211468) ne64a03.s1"
     gene            complement(22780. .29226)
                     /gene="NPY"
     CDS             complement(join(22780. .22804,24888. .24968,29039. .29226))
                     /gene="NPY"
                     /note="match to protein P01303 (PID:g128117);
                     H_RG118E13.1"
                     /codon_start=1
                     /product="neuropeptide Y precursor"
                     /protein_id="AAC08287.1"
                     /db_xref="GI:2992498"
                     /translation="MLGNKRLGLSGLTLALSLLVCLGALAEAYPSKPDNPGEDAPAED
                     MARYYSALRHYINLITRQRYGKRSSPETLISDLLMRESTENVPRTRLEDPAMW"
     repeat_region   23527. .23690
                     /rpt_family="Alu"
     misc_feature    24859. .24970
                     /note="match to EST AA662702 (NID:g2616693) nv06h11.s1"
     misc_feature    complement(24859. .24970)
                     /gene="NPY"
                     /note="match to EST AA226171 (NID:g1847487) nc11g07.r1"
     misc_feature    24859. .24965
                     /note="match to EST AA502194 (NID:g2237161) ng73e12.s1"
     repeat_region   26493. .26667
                     /rpt_family="MER1_type"
     repeat_region   26686. .26872
                     /rpt_family="MER1_type"
     repeat_region   28886. .28964
                     /rpt_family="L2"
     misc_feature    29006. .30527
                     /note="CpG_island (%GC=65.4, o/e=0.80, #CpGs=123)"
     misc_feature    29037. .29223
                     /note="match to EST AA662702 (NID:g2616693) nv06h11.s1"
     repeat_region   32483. .33035
                     /rpt_family="L1"
     repeat_region   33034. .33965
                     /rpt_family="L1"
     repeat_region   34024. .34394
                     /rpt_family="MaLR"
     repeat_region   34395. .34584
                     /rpt_family="L1"
     repeat_region   35069. .35311
                     /rpt_family="L2"
     repeat_region   35345. .41477
                     /rpt_family="L1"
     repeat_region   41729. .42145
                     /rpt_family="MaLR"
     repeat_region   42234. .42446
                     /rpt_family="MER1_type"
     repeat_region   42807. .42990
                     /rpt_family="L2"
     repeat_region   43879. .44184
                     /rpt_family="Alu"
     repeat_region   44664. .44934
                     /rpt_family="Alu"
     repeat_region   45003. .45251
                     /rpt_family="L1"
     repeat_region   45338. .45585
                     /rpt_family="L2"
     repeat_region   45623. .46229
                     /rpt_family="L2"
     repeat_region   46230. .46277
                     /rpt_family="MER73-group"
     repeat_region   46278. .46504
                     /rpt_family="L2"
     repeat_region   46517. .46674
                     /rpt_family="L1"
     repeat_region   46715. .46852
                     /rpt_family="MER1_type"
     repeat_region   47387. .47478
                     /rpt_family="L1"
     repeat_region   47824. .48086

  Query Match             12.8%;  Score 33.2;  DB 9;  Length 185516;
  Best Local Similarity   36.4%;  Pred. No. 8.8;
  Matches   43;  Conservative   22;  Mismatches   53;  Indels    0;  Gaps    0;

Qy      60 gacgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugug 119
           || |:   ::|:  |:|  |||| | || |    |||     : |    :| : | :|:|
Db 111357 GAGGTGGTTTCTCCCTCCCCCACCAGAAACTGGCCGGAAAACTTACTCCTCATCGATGTG 111416

Qy     120 ugggaagcagucaguauaauucccgucgugugguggugacgccucacgacguauuuguc 177
           :| | ||  |:||   :| ::  | : |:|:   |:| | |   | |||: : :| |
Db 111417 TGAGCAGTGGTCAAAGTATTTGTCATGGTGTAATGTGTCACAGAATGACTTTTATGGC 111474


RESULT  12
AL450993
LOCUS       AL450993   289563 bp    DNA             HTG       04-JUL-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-212D5, *** SEQUENCING IN
            PROGRESS ***, 39 unordered pieces.
ACCESSION   AL450993 AC025357
VERSION     AL450993.7  GI:14625625
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 289563)
  AUTHORS   Mclay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 6, 2001 this sequence version replaced gi:14588715.
            Draft Sequence Produced by Whitehead Institute/MIT Center for
            Genome Research, 320 Charles Street,
            Cambridge, MA 02141, USA
            http://www-seq.wi.mit.edu
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA212D5
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: M13; M77815; 44% of reads
            Sequencing vector: plasmid; L08752; 55% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 265164 bases at least Q40
            Consensus quality: 275491 bases at least Q30
            Consensus quality: 281753 bases at least Q20
            Insert size: 285763; sum-of-contigs
            Insert size: 170183; 6.7% error; agarose-fp
            Quality coverage: 3.27x in Q20 bases; sum-of-contigs Quality
            coverage: 6.17x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 39 contigs. The true order of the pieces
```

```
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     9307: contig of 9307 bp in length
*     9308  9407: gap of      100 bp
*     9408    13091: contig of 3684 bp in length
*    13092 13191: gap of      100 bp
*    13192    16963: contig of 3772 bp in length
*    16964 17063: gap of      100 bp
*    17064    28043: contig of 10980 bp in length
*    28044 28143: gap of      100 bp
*    28144    32213: contig of 4070 bp in length
*    32214 32313: gap of      100 bp
*    32314    39600: contig of 7287 bp in length
*    39601 39700: gap of      100 bp
*    39701    41803: contig of 2103 bp in length
*    41804 41903: gap of      100 bp
*    41904    47480: contig of 5577 bp in length
*    47481 47580: gap of      100 bp
*    47581    50607: contig of 3027 bp in length
*    50608 50707: gap of      100 bp
*    50708    65499: contig of 14792 bp in length
*    65500 65599: gap of      100 bp
*    65600    70500: contig of 4901 bp in length
*    70501 70600: gap of      100 bp
*    70601    78801: contig of 8201 bp in length
*    78802 78901: gap of      100 bp
*    78902    88610: contig of 9709 bp in length
*    88611 88710: gap of      100 bp
*    88711    91276: contig of 2566 bp in length
*    91277 91376: gap of      100 bp
*    91377   103516: contig of 12140 bp in length
*   103517 103616: gap of      100 bp
*   103617   106242: contig of 2626 bp in length
*   106243 106342: gap of      100 bp
*   106343   114051: contig of 7709 bp in length
*   114052 114151: gap of      100 bp
*   114152   128614: contig of 14463 bp in length
*   128615 128714: gap of      100 bp
*   128715   134258: contig of 5544 bp in length
*   134259 134358: gap of      100 bp
*   134359   139476: contig of 5118 bp in length
*   139477 139576: gap of      100 bp
*   139577   149866: contig of 10290 bp in length
*   149867 149966: gap of      100 bp
*   149967   159641: contig of 9675 bp in length
*   159642 159741: gap of      100 bp
*   159742   173858: contig of 14117 bp in length
*   173859 173958: gap of      100 bp
*   173959   181673: contig of 7715 bp in length
*   181674 181773: gap of      100 bp
*   181774   186169: contig of 4396 bp in length
*   186170 186269: gap of      100 bp
*   186270   201406: contig of 15137 bp in length
*   201407 201506: gap of      100 bp
*   201507   203507: contig of 2001 bp in length
*   203508 203607: gap of      100 bp
*   203608   227405: contig of 23798 bp in length
*   227406 227505: gap of      100 bp
*   227506   230659: contig of 3154 bp in length
*   230660 230759: gap of      100 bp
*   230760   233331: contig of 2572 bp in length
*   233332 233431: gap of      100 bp
*   233432   236751: contig of 3320 bp in length
*   236752 236851: gap of      100 bp
*   236852   242617: contig of 5766 bp in length
*   242618 242717: gap of      100 bp
*   242718   245181: contig of 2464 bp in length
*   245182 245281: gap of      100 bp
*   245282   257886: contig of 12605 bp in length
*   257887 257986: gap of      100 bp
*   257987   265713: contig of 7727 bp in length
*   265714 265813: gap of      100 bp
*   265814   271565: contig of 5752 bp in length
*   271566 271665: gap of      100 bp
*   271666   276811: contig of 5146 bp in length
*   276812 276911: gap of      100 bp
*   276912   281191: contig of 4280 bp in length
*   281192 281291: gap of      100 bp
*   281292   289563: contig of 8272 bp in length.
FEATURES             Location/Qualifiers
     source          1. .289563
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-212D5"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .9307
                     /note="assembly_fragment:00169
                     fragment_chain:1"
     misc_feature    9408. .13091
                     /note="assembly_fragment:00741
                     fragment_chain:1"
     misc_feature    13192. .16963
                     /note="assembly_fragment:01915
                     fragment_chain:1"
     misc_feature    17064. .28043
                     /note="assembly_fragment:00207
                     fragment_chain:2"
     misc_feature    28144. .32213
                     /note="assembly_fragment:00775
                     fragment_chain:2"
     misc_feature    32314. .39600
                     /note="assembly_fragment:00306
                     fragment_chain:2"
     misc_feature    39701. .41803
                     /note="assembly_fragment:00535
                     fragment_chain:3"
     misc_feature    41904. .47480
                     /note="assembly_fragment:00988
                     fragment_chain:3"
     misc_feature    47581. .50607
                     /note="assembly_fragment:01502
                     fragment_chain:4"
     misc_feature    50708. .65499
                     /note="assembly_fragment:00022
                     fragment_chain:4"
     misc_feature    65600. .70500
                     /note="assembly_fragment:01705
                     fragment_chain:5"
     misc_feature    70601. .78801
                     /note="assembly_fragment:02128
                     fragment_chain:5"
     misc_feature    78902. .88610
                     /note="assembly_fragment:01837
                     fragment_chain:6"
     misc_feature    88711. .91276
                     /note="assembly_fragment:00450
                     fragment_chain:6"
     misc_feature    91377. .103516
                     /note="assembly_fragment:00080"
     misc_feature    103617. .106242
                     /note="assembly_fragment:00130"
     misc_feature    106343. .114051
                     /note="assembly_fragment:00446"
     misc_feature    114152. .128614
                     /note="assembly_fragment:00448"
     misc_feature    128715. .134258
                     /note="assembly_fragment:00458"
     misc_feature    134359. .139476
                     /note="assembly_fragment:00731"
     misc_feature    139577. .149866
                     /note="assembly_fragment:01002"
```

```
     misc_feature    149967. .159641
                     /note="assembly_fragment:01439"
     misc_feature    159742. .173858
                     /note="assembly_fragment:01464"
     misc_feature    173959. .181673
                     /note="assembly_fragment:01538"
     misc_feature    181774. .186169
                     /note="assembly_fragment:01634"
     misc_feature    186270. .201406
                     /note="assembly_fragment:01661"
     misc_feature    201507. .203507

  Query Match             12.8%;  Score 33.2;  DB 2;  Length 289563;
  Best Local Similarity   36.4%;  Pred. No. 8.1;
  Matches   43;  Conservative   22;  Mismatches   53;  Indels    0;  Gaps    0;

Qy      60 gacgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugug 119
           || |:   ::|:  |:|  |||| | || |    |||     : |    :| : | :|:|
Db 150648 GAGGTGGTTTCTCCCTCCCCCACCAGAAACTGGCCGGAAAACTTACTCCTCATCGATGTG 150707

Qy     120 ugggaagcagucaguauaauucccgucgugugguggugacgccucacgacguauuuguc 177
           :| | ||  |:||   :| ::  | : |:|:   |:| | |   | ||| : : :| |
Db 150708 TGAGCAGTGGTCAAAGTATTTGTCATGGTGTAATGTGTCACAGAATGACTTTTATGGC 150765


RESULT  13
CNS07EGP/c
LOCUS       CNS07EGP   172751 bp    DNA             ROD       30-MAY-2001
DEFINITION  BAC 13C18 of library CITB 29-Sv from chromosome 10 of strain 129 of
            Mus musculus.
ACCESSION   AL591826
VERSION     AL591826.1  GI:14275545
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 172751)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is telomeric to the Ifn-gamma gene.
FEATURES             Location/Qualifiers
     source          1. .172751
                     /organism="Mus musculus"
                     /strain="129"
                     /db_xref="taxon:10090"
                     /chromosome="10"
                     /clone="13C18"
                     /clone_lib="CITB 129/Sv"
                     /note="cosegregates with D10Mit74, D10Mit14 and D10Mit234"
     repeat_region   1113. .1238
                     /rpt_family="D10Pas11:(TG)n"
     repeat_region   1238. .1275
                     /rpt_family="(GA)n"
     misc_feature    1939. .2055
                     /note="Region conserved with human (AC005294)"
     repeat_region   2150. .2228
                     /rpt_family="SINE/ID"
     repeat_region   2234. .2398
                     /rpt_family="GA-rich"
     repeat_region   2559. .2678
                     /rpt_family="LINE/L1"
     repeat_region   2795. .2829
                     /rpt_family="(TG)n"
     repeat_region   2874. .2954
                     /rpt_family="SINE/B2"
     repeat_region   3148. .3212
                     /rpt_family="(TA)n"
     repeat_region   3213. .3251
                     /rpt_family="(CA)n"
     repeat_region   3338. .3465
                     /rpt_family="LINE/L2"
     repeat_region   3941. .4089
                     /rpt_family="LINE/L2"
     repeat_region   4332. .4410
                     /rpt_family="LINE/L2"
     repeat_region   4417. .4560
                     /rpt_family="SINE/B2"
     repeat_region   5509. .5584
                     /rpt_family="(CAAAA)n"
     misc_feature    5600. .5727
                     /note="Region containing AA162222"
     repeat_region   5807. .5851
                     /rpt_family="LTR/MaLR"
     repeat_region   5820. .6050
                     /rpt_family="LTR/MaLR"
     repeat_region   6051. .6243
                     /rpt_family="SINE/B2"
     repeat_region   6244. .6407
                     /rpt_family="LTR/MaLR"
     misc_feature    7282. .7305
                     /note="identical to 7406. .7429"
     misc_feature    7406. .7429
                     /note="identical to 7282. .7305"
     misc_feature    7508. .7535
                     /note="identical to 7554. .7581"
     misc_feature    7554. .7581
                     /note="identical to 7508. .7535"
     repeat_region   9065. .9256
                     /rpt_family="LTR"
     repeat_region   9257. .9355
                     /rpt_family="(TC)n"
     repeat_region   9356. .9564
                     /rpt_family="LTR"
     repeat_region   9697. .9929
                     /rpt_family="LINE/L1"
     repeat_region   10016. .10067
                     /rpt_family="(CCA)n"
     misc_feature    10540. .10803
                     /note="Region conserved with human (AC005294)"
     repeat_region   10808. .10835
                     /rpt_family="AT_rich"
     repeat_region   11225. .11637
                     /rpt_family="LTR/MaLR"
     repeat_region   12879. .13008
                     /rpt_family="SINE/Alu"
     repeat_region   12912. .13019
                     /rpt_family="SINE/B4"
     misc_feature    13039. .13068
                     /note="identical to 13067. .13096"
     misc_feature    13067. .13096
                     /note="identical to 13039. .13068"
     repeat_region   13193. .13240
                     /rpt_family="LTR/Retroviral"
     repeat_region   13241. .13312
                     /rpt_family="(TG)n"
     repeat_region   13313. .13402
                     /rpt_family="LTR/Retroviral"
     repeat_region   14077. .14181
                     /rpt_family="LINE/L1"
     repeat_region   14182. .14223
                     /rpt_family="(CTA)n"
     repeat_region   14224. .14404
                     /rpt_family="LINE/L1"
     misc_feature    16814. .16975
                     /note="Region conserved with human (AC005294)"
     repeat_region   18915. .19059
                     /rpt_family="SINE/Alu"
     repeat_region   19063. .19101
                     /rpt_family="(CAAA)n"
     repeat_region   19317. .19752
                     /rpt_family="LINE/L1"
     repeat_region   20508. .20552
```

```
                /rpt_family="(TATG)n"
repeat_region   20603. .20840
                /rpt_family="SINE/B4"
repeat_region   20917. .21270
                /rpt_family="(TCTA)n"
repeat_region   22282. .22511
                /rpt_family="SINE/MIR"
repeat_region   22576. .22739
                /rpt_family="LTR/MaLR"
repeat_region   22916. .23117
                /rpt_family="SINE/B4"
repeat_region   23118. .23172
                /rpt_family="(TG)n"
repeat_region   23173. .23242
                /rpt_family="SINE/B4"
repeat_region   23246. .23302
                /rpt_family="(CATATA)n"
repeat_region   23877. .23904
                /rpt_family="(CTG)n"
repeat_region   25454. .25634
                /rpt_family="(CA)n"
repeat_region   25668. .25848
                /rpt_family="(CA)n"
repeat_region   25872. .26151
                /rpt_family="(CA)n"
misc_feature    27355. .27616
                /note="Region conserved with human ( AC005294)"
repeat_region   28102. .28405
                /rpt_family="LTR/MaLR"
repeat_region   28416. .28499
                /rpt_family="LINE/L1"
repeat_region   28773. .28800
                /rpt_family="T-rich"
repeat_region   28848. .28965
                /rpt_family="LINE/L1"
repeat_region   29061. .29191
                /rpt_family="(TCTA)n"
repeat_region   29195. .29261
                /rpt_family="SINE/ID"
repeat_region   29483. .29783
                /rpt_family="LINE/L1"
repeat_region   29784. .29926
                /rpt_family="SINE/Alu"
repeat_region   29927. .30343
                /rpt_family="LINE/L1"
repeat_region   30344. .30386
                /rpt_family="D10Pas8:(TG)n"
repeat_region   30387. .30451
                /rpt_family="LINE/L1"
repeat_region   30750. .31139
                /rpt_family="LTR/MaLR"
repeat_region   31913. .32039
                /rpt_family="(CATATA)n"
misc_feature    32168. .32367
                /note="Region conserved with human (AC005294)"
repeat_region   32762. .32953
                /rpt_family="LTR/Retroviral"
repeat_region   33037. .33305
                /rpt_family="(GA)n"
repeat_region   33466. .33835
                /rpt_family="LTR/MaLR"
repeat_region   34171. .34265
                /rpt_family="SINE/Alu"
misc_feature    34309. .34346
                /note="identical to 34340. .34377"
misc_feature    34340. .34377
                /note="identical to 34309. .34346"
repeat_region   34445. .34509
                /rpt_family="D10Mit74:(TG)n"
misc_feature    36750. .37360
                /note="Region conserved with human (AC005294)"
misc_feature    37047. .37124
                /note="Region containing AA162222"
repeat_region   37754. .37783
                /rpt_family="AT_rich"
repeat_region   37831. .37933
                /rpt_family="SINE/B4"
repeat_region   39191. .39213
                /rpt_family="(CAAA)n"
repeat_region   39265. .39406
                /rpt_family="SINE/Alu"

  Query Match             12.6%;  Score 32.6;  DB 10;  Length 172751;
  Best Local Similarity   40.2%;  Pred. No. 14;
  Matches   35;  Conservative   18;  Mismatches   34;  Indels    0;  Gaps    0;

Qy    145 ucgugugguggugacgccucacgacguauuuguccgcugugcagagcguaguaccaagggc 204
          : |:|:| | : | |  :| | |||  : :|:|| |:|   || || : |:| || ||
Db  11050 TTGTGTGGGCTCAGGGTTCTCTACGGCTATGTCCACTGCAGAGTGCATTGTAGCATGGCG 10991

Qy    205 ugcaccccgguuuuuguuccaagcgga 231
          :   |   ||::: : ::| | || ||
Db  10990 TTAGCATTGGTTTGTCTTCTAGGCAGA 10964


RESULT  14
PPM76VP1
LOCUS       PPM76VP1     2354 bp    DNA             VRL       17-FEB-1998
DEFINITION  Monkey B-lymphotropic papovavirus genes for VP-1, 2, 3 and large T
            antigen, complete and partial cds, strain LPV-76.
ACCESSION   D00056 N00056
VERSION     D00056.1  GI:221987
KEYWORDS    VP-1; VP-2; VP-3; large T antigen.
SOURCE      Monkey B-lymphotropic papovavirus (strain:LPV-76) DNA, clone:pL76.
  ORGANISM  Monkey B-lymphotropic papovavirus
            Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
            Papillomavirus.
REFERENCE   1  (sites)
  AUTHORS   Pawlita,M., Clad,A. and zur Hausen,H.
  TITLE     Complete DNA sequence of lymphotropic papovavirus: prototype of a
            new species of the polyomavirus genus
  JOURNAL   Virology 143 (1), 196-211 (1985)
  MEDLINE   86045879
REFERENCE   2  (bases 1 to 2354)
  AUTHORS   Kanda,T., Furuno,A. and Yoshiike,K.
  TITLE     Mutation in the VP-1 gene is responsible for the extended host
            range of a monkey B-lymphotropic papovavirus mutant capable of
            growing in T-lymphoblastoid cells
  JOURNAL   J. Virol. 59 (2), 531-534 (1986)
  MEDLINE   86281860
FEATURES             Location/Qualifiers
     source          1. .2354
                     /organism="Monkey B-lymphotropic papovavirus"
                     /strain="LPV-76"
                     /db_xref="taxon:10574"
                     /clone="pL76"
                     /map="PstI B segment"
                     /note="mutant"
     CDS             <1. .759
                     /codon_start=1
                     /product="VP-2"
                     /protein_id="BAA00041.1"
                     /db_xref="GI:441186"
                     /translation="GVTTFGYSKEVPVVNMALVPWFPQVDYLFPGFTSFSYYLNAVLD
                     WGESLFHAVGREVWRHLMRQATLQIGQATRAVAVRSTNELSHTLAQIAENARWALTSG
                     PVHIYSSVQDYYRYLPARNPIQLRQEYRNRGEPPPSRADFEYQENREGQRARRELGYD
                     EPRSGQYVEHYTAPGGAHQRVTQDWMLPLILGLYGDITPTWEVELNKLEKEEDGPSKK
                     KARRSMQKNMPYSRSRPQAPSKRRSRGARSKNRA"
     CDS             46. .759
                     /codon_start=1
                     /product="VP-3"
                     /protein_id="BAA00042.1"
                     /db_xref="GI:441187"
                     /translation="MALVPWFPQVDYLFPGFTSFSYYLNAVLDWGESLFHAVGREVWR
                     HLMRQATLQIGQATRAVAVRSTNELSHTLAQIAENARWALTSGPVHIYSSVQDYYRYL
                     PARNPIQLRQEYRNRGEPPPSRADFEYQENREGQRARRELGYDEPRSGQYVEHYTAPG
```

```
                     GAHQRVTQDWMLPLILGLYGDITPTWEVELNKLEKEEDGPSKKKARRSMQKNMPYSRS
                     RPQAPSKRRSRGARSKNRA"
     CDS             638. .1744
                     /codon_start=1
                     /product="VP-1"
                     /protein_id="BAA00043.1"
                     /db_xref="GI:441188"
                     /translation="MAPQRKRQDGACKKTCPIPAPVPRLLVKGGVEVLEVRTGPDAIT
                     QIEAYLNPRMGNNIPSEDLYGYSNSINTASSKASDTPNKGTLPCYSVAVIKLPLLNED
                     MTCDTILMWEAVSVKTEVVGISSLVNLHQGGKYIYGSYSGCVPVQGTTYHMFAVGGEP
                     LELQGLVASSTATYPDDVVAIKNMKPGNQGLDPKAKALLDKDGKYPVEVWCPDPSKNE
                     NTRYYGSFTGGATTPPVMQFTNSVTTVLLDENGVGPLCKGDKLFLSCADIAGVHTNYS
                     ETQNWRGLPRYFNVTLRKRIVKNPYPVSSLLNSFFSGLMPQIQGQPMEGVSGQVEEVR
                     IFEGTEGLPGDPDLNRYVDKFCQNQTVLPVSNDM"
     variation       858
                     /note="C in LPV-76 changed to T in LPV-02, and in LPV-K38
                     in [1]"
                     /replace="t"
     variation       888
                     /note="G in LPV-76 changed to A in LPV-K38 in [1]"
                     /replace="a"
     variation       1056
                     /note="A in LPV-76 changed to C in LPV-02"
                     /replace="c"
     variation       1475
                     /note="A in LPV-76 changed to G in LPV-K38 in [1]"
                     /replace="g"
     variation       1476
                     /note="A in LPV-76 changed to G in LPV-02, and to T in
                     LPV-K38 in [1]"
                     /replace="g or t"
     variation       1709
                     /note="A in LPV-76 changed to C in LPV-K38 in [1]"
                     /replace="c"
     variation       1729
                     /note="A in LPV-76 changed to C in LPV-02"
                     /replace="c"
     variation       1756
                     /note="T in LPV-76 changed to G in LPV-K38 in [1]"
                     /replace="g"
     variation       1806
                     /note="G in LPV-76 changed to A in LPV-02"
                     /replace="a"
     CDS             complement(1808. .>2354)
                     /codon_start=2
                     /product="large T antigen"
                     /protein_id="BAA00044.1"
                     /db_xref="GI:221988"
                     /translation="ADKLSFELGCAIDQFCVLLDDVKGQITLNKHLQPGQGVNNLDNL
                     RDHLDGTIKVNLEKKHVNKRSQIFPPVIMTMNEYLLPPTIGVRFALHLHLKPKAYLKQ
                     SLEKSDLVAKRILNSGYTILLLLLWYNPVDSFTPKVQEKVVQWKETLEKYVSITQFGN
                     IQQNIIDGKDPLHGIVIEEQM"
BASE COUNT      705 a    470 c    507 g    672 t
ORIGIN


  Query Match             12.5%;  Score 32.4;  DB 14;  Length 2354;
  Best Local Similarity   42.6%;  Pred. No. 35;
  Matches   52;  Conservative   14;  Mismatches   56;  Indels    0;  Gaps    0;

Qy     118 ugugggaagcagucaguauaauucccgucgugugugugacgccucacgacguauuuguc 177
           :|:|||||||||:   : :||   | |  |:   :||     ||:||| |  :| ::  |
Db     966 TGTGGGAAGCAGTGTCTGTAAAGACTGAAGTTGTTGGAATTTCCTCACTAGTTAATTTGC 1025

Qy     178 cgcugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaa 237
             | |  ||  |:|   | | |||    || | || : :    ||  || | |  | |
Db    1026 ACCAGGGAGGAAAGTACATCTATGGGTCATACTCAGGGTGTGTCCCCGTGCAGGGCACTA 1085

Qy     238 cc 239
           ||
Db    1086 CC 1087
```

```
RESULT  15
PPMVP1M
LOCUS       PPMVP1M      2354 bp    DNA   circular  VRL      03-AUG-1993
DEFINITION  Monkey B-lymphotropic papovavirus mutant (LPV-76) PstI B fragment
            encoding VP1, VP2, VP3 and T-antigen.
ACCESSION   M14494
VERSION     M14494.1  GI:333288
KEYWORDS    T-antigen; antigen; capsid protein; large T-antigen; viral capsid
            protein vp1; viral capsid protein vp2; viral capsid protein vp3.
SOURCE      B-Lymphotropic papovavirus (strain LPV-76) DNA grown in African
            green monkey T-lymphotropic cell line, clone pl76.
  ORGANISM  Monkey B-lymphotropic papovavirus
            Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
            Papillomavirus.
REFERENCE   1  (bases 1 to 2354)
  AUTHORS   Kanda,T., Furuno,A. and Yoshiike,K.
  TITLE     Mutation in the VP-1 Gene Is Responsible for the Extended Host
            Range of a Monkey B-Lymphotropic Papovavirus Mutant Capable of
            Growing in T-Lymphoblastoid Cells
  JOURNAL   J. Virol. 59, 531-534 (1986)
  MEDLINE   86281860
COMMENT     Draft entry and clean copy of sequence [1] kindly provided by
            T.Kanda, 08-DEC-1986.
FEATURES             Location/Qualifiers
     source          1. .2354
                     /organism="Monkey B-lymphotropic papovavirus"
                     /db_xref="taxon:10574"
     CDS             <1. .759
                     /note="B-lymphotropic papovavirus viral protein 2"
                     /codon_start=1
                     /protein_id="AAA47074.1"
                     /db_xref="GI:333289"
                     /translation="GVTTFGYSKEVPVVNMALVPWFPQVDYLFPGFTSFSYYLNAVLD
                     WGESLFHAVGREVWRHLMRQATLQIGQATRAVAVRSTNELSHTLAQIAENARWALTSG
                     PVHIYSSVQDYYRYLPARNPIQLRQEYRNRGEPPPSRADFEYQENREGQRARRELGYD
                     EPRSGQYVEHYTAPGGAHQRVTQDWMLPLILGLYGDITPTWEVELNKLEKEEDGPSKK
                     KARRSMQKNMPYSRSRPQAPSKRRSRGARSKNRA"
     CDS             46. .759
                     /note="B-lymphotropic papovavirus viral protein 3"
                     /codon_start=1
                     /protein_id="AAA47075.1"
                     /db_xref="GI:333290"
                     /translation="MALVPWFPQVDYLFPGFTSFSYYLNAVLDWGESLFHAVGREVWR
                     HLMRQATLQIGQATRAVAVRSTNELSHTLAQIAENARWALTSGPVHIYSSVQDYYRYL
                     PARNPIQLRQEYRNRGEPPPSRADFEYQENREGQRARRELGYDEPRSGQYVEHYTAPG
                     GAHQRVTQDWMLPLILGLYGDITPTWEVELNKLEKEEDGPSKKKARRSMQKNMPYSRS
                     RPQAPSKRRSRGARSKNRA"
     CDS             638. .1744
                     /note="B-lymphotropic papovavirus viral protein 1"
                     /codon_start=1
                     /protein_id="AAA47076.1"
                     /db_xref="GI:333291"
                     /translation="MAPQRKRQDGACKKTCPIPAPVPRLLVKGGVEVLEVRTGPDAIT
                     QIEAYLNPRMGNNIPSEDLYGYSNSINTASSKASDTPNKGTLPCYSVAVIKLPLLNED
                     MTCDTILMWEAVSVKTEVVGISSLVNLHQGGKYIYGSYSGCVPVQGTTYHMFAVGGEP
                     LELQGLVASSTATYPDDVVAIKNMKPGNQGLDPKAKALLDKDGKYPVEVWCPDPSKNE
                     NTRYYGSFTGGATTPPVMQFTNSVTTVLLDENGVGPLCKGDKLFLSCADIAGVHTNYS
                     ETQNWRGLPRYFNVTLRKRIVKNPYPVSSLLNSFFSGLMPQIQGQPMEGVSGQVEEVR
                     IFEGTEGLPGDPDLNRYVDKFCQNQTVLPVSNDM"
     gene            complement(1808. .2354)
                     /gene="T-antigen"
     CDS             complement(1808. .>2354)
                     /gene="T-antigen"
                     /codon_start=2
                     /product="T-antigen"
                     /protein_id="AAA47077.1"
                     /db_xref="GI:333292"
                     /translation="ADKLSFELGCAIDQFCVLLDDVKGQITLNKHLQPGQGVNNLDNL
                     RDHLDGTIKVNLEKKHVNKRSQIFPPVIMTMNEYLLPPTIGVRFALHLHLKPKAYLKQ
                     SLEKSDLVAKRILNSGYTILLLLLWYNPVDSFTPKVQEKVVQWKETLEKYVSITQFGN
                     IQQNIIDGKDPLHGIVIEEQM"
BASE COUNT      705 a    470 c    507 g    672 t
ORIGIN      At PstI site.
```

Query Match 12.5%; Score 32.4; DB 14; Length 2354;
Best Local Similarity 42.6%; Pred. No. 35;
Matches 52; Conservative 14; Mismatches 56; Indels 0; Gaps 0;

```
Qy     118 ugugggaagcagucaguauaauucccgucgugugguggugacgccucacgacguauuuguc 177
           :|:|||||||||:   : :||   | |  |:   :||     ||:||| |  :| ::  |
Db     966 TGTGGGAAGCAGTGTCTGTAAAGACTGAAGTTGTTGGAATTTCCTCACTAGTTAATTTGC 1025

Qy     178 cgcugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaa 237
             | | |  ||  |:|   | | |||    || | || : :    ||  || | |  | |
Db    1026 ACCAGGGAGGAAAGTACATCTATGGGTCATACTCAGGGTGTGTCCCCGTGCAGGGCACTA 1085

Qy     238 cc 239
           ||
Db    1086 CC 1087
```

Search completed: February 16, 2002, 16:15:04
Job time: 6537 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: February 16, 2002, 14:28:12 ; Search time 166.92 Seconds
(without alignments)
1330.261 Million cell updates/sec

Title: US-09-587-653-1
Perfect score: 259
Sequence: 1 aguuuggcgaccauggugga.....cccgcuuggaauuaaaaacu 259

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	259	100.0	9399	22	AAF23484	GBV-B virus genome
	2	257.4	99.4	259	22	AAC91998	GBV-B 3'X DNA sequ
	3	257.4	99.4	309	22	AAC92010	GBV-B 3'X sequence
	4	245.4	94.7	258	22	AAC91996	GBV-B 3'X RNA sequ
	5	82	31.7	82	22	AAC92012	GBV-B 3'X RNA sequ
c	6	32	12.4	3179	22	AAD09560	Human transporter
c	7	30.4	11.7	522	21	AAC81741	Human secreted pro
c	8	30.4	11.7	2342	21	AAA93110	Human secreted pro
	9	30	11.6	49999	20	AAZ23900	Human LOBO homolog
c	10	29.8	11.5	24183	22	AAS21771	Human gene for col
c	11	29.4	11.4	1933	22	AAH20379	Human RNA-binding
c	12	29.4	11.4	1993	22	AAH15309	Human cDNA sequenc
c	13	29.4	11.4	2092	21	AAA49680	Human quaking spli
c	14	29.4	11.4	2472	21	AAA49681	Human quaking spli
	15	29	11.2	775	21	AAC39642	Arabidopsis thalia
c	16	28.8	11.1	400	22	AAH97854	Murine 7-transmemb
	17	28.6	11.0	319608	21	AAH51601	Human chromosome 1
	18	28.6	11.0	319608	22	AAS09301	Human schizophreni
c	19	28.4	11.0	531	22	AAH97850	Murine 7-transmemb
c	20	28.4	11.0	560	22	AAH97844	Murine 7-transmemb
c	21	28.4	11.0	1548	22	AAF93854	Human cDNA encodin
	22	28.2	10.9	6623	19	AAV19051	Mitochondrial gene
	23	28.2	10.9	6623	19	AAV19053	Mitochondrial gene
	24	28	10.8	1723	13	AAQ28737	Human osteogenic p
	25	28	10.8	1723	14	AAQ53155	Sequence encoding
	26	28	10.8	1723	14	AAQ38946	Human osteogenic p
	27	28	10.8	1723	14	AAQ38735	Human OP-2. Homo
	28	28	10.8	1723	14	AAQ38859	Morphogen hOP2 cod
	29	28	10.8	1723	15	AAQ56200	hOP2 cDNA. Homo s
	30	28	10.8	1723	15	AAQ56233	hOP2 cDNA. Homo s
	31	28	10.8	1723	15	AAQ57917	Human osteogenic p
	32	28	10.8	1723	15	AAQ67313	Human OP-2. Homo
	33	28	10.8	1723	15	AAQ45118	Human OP-2. Homo
	34	28	10.8	1723	15	AAQ65393	Osteogenic protein
	35	28	10.8	1723	15	AAQ45164	Human OP-2. Homo
	36	28	10.8	1723	15	AAQ71426	Osteogenic protein
	37	28	10.8	1723	15	AAQ72706	hOP2-PP prepro for
	38	28	10.8	1723	17	AAT33443	Human osteogenic p
	39	28	10.8	1723	17	AAT02600	hOP-2 cDNA. Homo
	40	28	10.8	1723	18	AAV10347	Human OP-2 cDNA.
	41	28	10.8	1723	18	AAT97881	cDNA encoding huma
	42	28	10.8	1723	18	AAT73208	Human hippocampal
	43	28	10.8	1723	19	AAV32585	Human osteogenic p
	44	28	10.8	1723	19	AAV19535	Human osteogenic p
	45	28	10.8	1723	22	AAC89690	Human tissue morph

ALIGNMENTS

RESULT 1
AAF23484
ID AAF23484 standard; DNA; 9399 BP.
XX
AC AAF23484;
XX
DT 21-MAR-2001 (first entry)
XX
DE GBV-B virus genome.
XX
KW GBV-B; hepatitis C virus; HCV; vaccine; ds.
XX
OS GB virus B.
XX
PN WO200075337-A1.
XX
PD 14-DEC-2000.
XX
PF 02-JUN-2000; 2000WO-US15293.
XX
PR 04-JUN-1999; 99US-0137694.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Bukh J, Yanagi M, Emerson SU, Purcell RH;
XX
DR WPI; 2001-091214/10.
XX
PT New infectious nucleic acids of the GB virus-B clone, useful for
PT indirectly studying the molecular properties of hepatitis C virus (HCV)
PT and in developing vaccines and therapeutics for HCV -
XX
PS Claim 3; Page 60-63; 96pp; English.
XX

```
CC   The present invention relates to GB virus-B. The nucleic acid molecules
CC   of the invention are useful for indirectly studying the molecular
CC   properties of hepatitis C virus (HCV). The infectious nucleic acid
CC   sequence of the GB virus-B clone and the HCV/GBV-B chimeras may be used
CC   in the development of vaccines and therapeutics for HCV.
XX
SQ   Sequence 9399 BP; 2125 A; 2326 C; 2430 G; 2518 T; 0 other;


  Query Match           100.0%;  Score 259;  DB 22;  Length 9399;
  Best Local Similarity  76.1%;  Pred. No. 1.4e-78;
  Matches  197;  Conservative   62;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccaugucugaaggggaug 60
          ||:::|||||||||:||:|||:||||||||:::||||:|||||||:||:|:|||||||:|
Db   9141 agtttggcgaccatggtggatcagaaccgtttcgggtgaagccatggtctgaaggggatg 9200

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          |||:|||::|:|||:||:|||||||||||||:|:||||:|||:|||||:||:|||:|:|:
Db   9201 acgtcccttctggctcatccacaaaaaccgtctcgggtgggtgaggagtcctggctgtgt 9260

Qy    121 gggaagcagucaguauaauuucccgucgugugguggugacgccucacgacguauuuguccgc 180
          |||||||||:|||:|:||::||||:||:|:|:||:|||||:|||||||:|:::|:||||
Db   9261 gggaagcagtcagtataattcccgtcgtgtgtggtgacgcctcacgacgtatttgtccgc 9320

Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          :|:|||||||||:||:|||||||||:|||||||||:::::|::|||||||||||||||||
Db   9321 tgtgcagagcgtagtaccaagggctgcaccccggtttttgttccaagcggagggcaaccc 9380

Qy    241 ccgcuuggaauuaaaaacu 259
          |||||::|||||::|||||:
Db   9381 ccgcttggaattaaaaact 9399


RESULT   2
AAC91998
ID   AAC91998 standard; DNA; 259 BP.
XX
AC   AAC91998;
XX
DT   20-MAR-2001  (first entry)
XX
DE   GBV-B 3'X DNA sequence.
XX
KW   GBV-B; hepatitis GB viral-B; hepatitis C virus activity; ds.
XX
OS   Hepatitis virus.
XX
PN   WO200073466-A1.
XX
PD   07-DEC-2000.
XX
PF   22-MAY-2000; 2000WO-EP04622.
XX
PR   27-MAY-1999;   99GB-0012432.
XX
PA   (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI   Traboni C;
XX
DR   WPI; 2001-061544/07.
XX
PT   Novel GB viral sequence useful in in vivo assays to identify agents
PT   which modulate hepatitis C virus activity  -
XX
PS   Claim 1; Page 14; 76pp; English.
XX
CC   The present sequence is an isolated hepatitis GB viral-B (GBV-B)
CC   polynucleotide. The hepatitis GB agent was first discovered by the
CC   inoculation of tamarins with serum from a patient, whose initials were
CC   GB, affected by acute hepatitis. The serum induced hepatitis in all
CC   inoculated tamarins. The present sequence comprises the 3' terminus of
CC   GBV-B, and confers infectivity in tamarins on otherwise non-infective
CC   GBV-B genome. This sequence is useful in in vivo assays to identify
CC   agents which modulate hepatitis C virus (HCV) activity.
XX
SQ   Sequence 259 BP; 53 A; 64 C; 81 G; 61 T; 0 other;


  Query Match           99.4%;  Score 257.4;  DB 22;  Length 259;
  Best Local Similarity  76.1%;  Pred. No. 1.3e-78;
  Matches  197;  Conservative   61;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccaugucugaaggggaug 60
          ||:::|||||||||:||:|||:||||||||:::||||:|||||||:||:|:|||||||:|
Db      1 agtttggcgaccatggtggatcagaaccgtttcgggtgaagccatggtctgaaggggatg 60

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          |||:|||::|:|||:||:|||||||||||||:|:||||:|||:|||||:||:|||:|:|:
Db     61 acgtcccttctggctcatccacaaaaaccgtctcgggtgggtgaggagtcctggctgtgt 120

Qy    121 gggaagcagucaguauaauuucccgucgugugguggugacgccucacgacguauuuguccgc 180
          |||||||||:|||:|:||::||||:||:|:|:||:|||||:|||||||:|:::|:||||
Db    121 gggaagcagtcagtataattcccgtcgtgtgtggtgacgcctcacgacgtacttgtccgc 180

Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          :|:|||||||||:||:|||||||||:|||||||||:::::|::|||||||||||||||||
Db    181 tgtgcagagcgtagtaccaagggctgcaccccggtttttgttccaagcggagggcaaccc 240

Qy    241 ccgcuuggaauuaaaaacu 259
          |||||::|||||::|||||:
Db    241 ccgcttggaattaaaaact 259


RESULT   3
AAC92010
ID   AAC92010 standard; DNA; 309 BP.
XX
AC   AAC92010;
XX
DT   20-MAR-2001  (first entry)
XX
DE   GBV-B 3'X sequence and 3'UTR region.
XX
KW   GBV-B; hepatitis GB viral-B; hepatitis C virus activity; ds.
XX
OS   Hepatitis virus.
XX
PN   WO200073466-A1.
XX
PD   07-DEC-2000.
XX
PF   22-MAY-2000; 2000WO-EP04622.
XX
PR   27-MAY-1999;   99GB-0012432.
XX
PA   (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI   Traboni C;
XX
DR   WPI; 2001-061544/07.
XX
PT   Novel GB viral sequence useful in in vivo assays to identify agents
PT   which modulate hepatitis C virus activity  -
XX
PS   Example 2; Fig 3; 76pp; English.
XX
CC   The present invention relates to a hepatitis GB viral-B (GBV-B)
CC   polynucleotide (see AAC91998 and AAC91996). The hepatitis GB agent was
CC   first  discovered by the inoculation of tamarins with serum from a
CC   patient,  whose initials were GB, affected by acute hepatitis. The serum
CC   induced  hepatitis in all inoculated tamarins. The present sequence
CC   comprises the  sequence in AAC91998 (the 3' terminus of GBV-B, plus the
CC   3' UTR. This  sequence confers infectivity in tamarins on otherwise
CC   non-infective  GBV-B genome. This sequence is useful in in vivo assays to
CC   identify  agents which modulate hepatitis C virus (HCV) activity.
```

```
XX
SQ   Sequence 309 BP; 64 A; 80 C; 100 G; 65 T; 0 other;


  Query Match             99.4%;  Score 257.4;  DB 22;  Length 309;
  Best Local Similarity   76.1%;  Pred. No. 1.4e-78;
  Matches  197;  Conservative   61;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
          ||:::|||||||:||:|||:|||||||||:::|||:|||||||:||:|:|||||||:|
Db     51 agtttggcgaccatggtggatcagaaccgtttcgggtgaagccatggtctgaaggggatg 110

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          |||:|||::|:|||:||:|||||||||||:|:||||:|||:||||||:||:|||:|:|:
Db    111 acgtcccttctggctcatccacaaaaaccgtctcgggtgggtgaggagtcctggctgtgt 170

Qy    121 gggaagcagucaguauaaauucccgucgugugugggugacgccucacgacguauuuguccgc 180
          |||||||||:|||:|:||::||||:||:|:|:|||:||||||:|||||||:| ::|:||||
Db    171 gggaagcagtcagtataattcccgtcgtgtgtggtgacgcctcacgacgtacttgtccgc 230

Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          :|:|||||||||:||:|||||||||:|||||||||:::::|::||||||||||||||||||
Db    231 tgtgcagagcgtagtaccaagggctgcaccccggtttttgttccaagcggagggcaaccc 290

Qy    241 ccgcuuggaauuaaaaacu 259
          |||||::|||||::|||||:
Db    291 ccgcttggaattaaaaact 309


RESULT   4
AAC91996
ID   AAC91996 standard; RNA; 258 BP.
XX
AC   AAC91996;
XX
DT   20-MAR-2001  (first entry)
XX
DE   GBV-B 3'X RNA sequence #1.
XX
KW   GBV-B; hepatitis GB viral-B; hepatitis C virus activity; ds.
XX
OS   Hepatitis virus.
XX
PN   WO200073466-A1.
XX
PD   07-DEC-2000.
XX
PF   22-MAY-2000; 2000WO-EP04622.
XX
PR   27-MAY-1999;  99GB-0012432.
XX
PA   (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI   Traboni C;
XX
DR   WPI; 2001-061544/07.
XX
PT   Novel GB viral sequence useful in in vivo assays to identify agents
PT   which modulate hepatitis C virus activity  -
XX
PS   Claim 1; Page 62; 76pp; English.
XX
CC   The present sequence is an isolated hepatitis GB viral-B (GBV-B)
CC   polynucleotide. The hepatitis GB agent was first discovered by the
CC   inoculation of tamarins with serum from a patient, whose initials were
CC   GB, affected by acute hepatitis. The serum induced hepatitis in all
CC   inoculated tamarins. The present sequence comprises the 3' terminus of
CC   GBV-B, and confers infectivity in tamarins on otherwise non-infective
CC   GBV-B genome. This sequence is useful in in vivo assays to identify
CC   agents which modulate hepatitis C virus (HCV) activity.
XX
SQ   Sequence 258 BP; 53 A; 64 C; 80 G; 61 U; 0 other;


  Query Match             94.7%;  Score 245.4;  DB 22;  Length 258;
  Best Local Similarity   99.2%;  Pred. No. 1.7e-74;
  Matches  257;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     61 acgucccuucuggcucauccacaaaaacc-ucucgggugggugaggaguccuggcugugu 119

Qy    121 gggaagcagucaguauaaauucccgucgugugugggugacgccucacgacguauuuguccgc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    120 gggaagcagucaguauaaauucccgucgugugugggugacgccucacgacguacuuguccgc 179

Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 239

Qy    241 ccgcuuggaauuaaaaacu 259
          |||||||||||||||||||
Db    240 ccgcuuggaauuaaaaacu 258


RESULT   5
AAC92012
ID   AAC92012 standard; RNA; 82 BP.
XX
AC   AAC92012;
XX
DT   20-MAR-2001  (first entry)
XX
DE   GBV-B 3'X RNA sequence #2.
XX
KW   GBV-B; hepatitis GB viral-B; ds.
XX
OS   Hepatitis virus.
XX
PN   WO200073466-A1.
XX
PD   07-DEC-2000.
XX
PF   22-MAY-2000; 2000WO-EP04622.
XX
PR   27-MAY-1999;  99GB-0012432.
XX
PA   (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI   Traboni C;
XX
DR   WPI; 2001-061544/07.
XX
PT   Novel GB viral sequence useful in in vivo assays to identify agents
PT   which modulate hepatitis C virus activity  -
XX
PS   Disclosure; Fig 4; 76pp; English.
XX
CC   The present sequence is an isolated hepatitis GB viral-B (GBV-B)
CC   polynucleotide. The hepatitis GB agent was first discovered by the
CC   inoculation of tamarins with serum from a patient, whose initials were
CC   GB, affected by acute hepatitis. The serum induced hepatitis in all
CC   inoculated tamarins. The present sequence comprises the 3' terminus of
CC   GBV-B.
XX
SQ   Sequence 82 BP; 19 A; 23 C; 23 G; 17 U; 0 other;


  Query Match             31.7%;  Score 82;  DB 22;  Length 82;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-18;
  Matches   82;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy    178 cgcugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaa 237
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 cgcugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaa 60

Qy    238 cccccgcuuggaauuaaaaacu 259
          |||||||||||||||||||||||
Db     61 cccccgcuuggaauuaaaaacu 82
```

RESULT 6
AAD09560/c
ID AAD09560 standard; cDNA; 3179 BP.
XX
AC AAD09560;
XX
DT 10-SEP-2001 (first entry)
XX
DE Human transporter and ion channel-9 (TRICH-9) cDNA.
XX
KW Human; transporter and ion channel-9; TRICH-9; vaccine; cystic fibrosis;
KW gene therapy; amyotrophic lateral sclerosis; amnesia; muscular dystrophy;
KW hypertension; angina; neurological disorder; asthma; bipolar disorder;
KW dementia; depression; Alzheimer's disease; epilepsy; mood; arrhythmia;
KW Pick's disease; ischaemic cerebrovascular disease; AIDS; anxiety; stroke;
KW Huntington's disease; Parkinson's disease; cerebral neoplasm; allergy;
KW demyelinating disease; mental disorder; Schizophrenia; polymyositis;
KW muscle disorder; cardiomyopathy; cataract; myocarditis; Grave's disease;
KW dermatomyositis; diabetes mellitus; immunological disorder; psoriasis;
KW rheumatoid arthritis; Sjogren's syndrome; systemic lupus erythematosus;
KW sickle cell anaemia; Wilson's disease; infertility; Cushing's disease;
KW scleroderma; pulmonary artery stenosis; nootropic; Addison's disease;
KW malabsorption syndrome; hypercholesterolaemia; cancer; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 396..1592
FT /*tag= a
FT /product= "Human TRICH-9 protein"
XX
PN WO200146258-A2.
XX
PD 28-JUN-2001.
XX
PF 22-DEC-2000; 2000WO-US35095.
XX
PR 23-DEC-1999; 99US-0172000.
PR 14-JAN-2000; 2000US-0176083.
PR 21-JAN-2000; 2000US-0177332.
PR 28-JAN-2000; 2000US-0178572.
PR 02-FEB-2000; 2000US-0179758.
PR 10-FEB-2000; 2000US-0181625.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Baughn MR, Burford N, Au-Young J, Lu DAM, Yang J, Reddy R, Lal P;
PI Hillman JL, Azimzai Y, Yue H, Nguyen DB, Yao MG, Gandhi AR;
PI Tang YT, Khan FA;
XX
DR WPI; 2001-418042/44.
DR P-PSDB; AAE04896.
XX
PT Novel human transporter and ion channel proteins useful for treating
PT and preventing transport, neurological, muscle and immunological
PT disorders -
XX
PS Claim 5; Page 146-147; 160pp; English.
XX
CC The present sequence is transporter and ion channel-9 (TRICH-9) cDNA.
CC TRICH is used as vaccine. TRICH is useful for treating a disease or
CC condition associated with decreased expression of functional TRICH,
CC such as transport disorder including amyotrophic lateral sclerosis,
CC cystic fibrosis, Becker's muscular dystrophy, Charcot-Marie Tooth
CC disease, Duchenne muscular dystrophy, angina and hypertension,
CC neurological disorders including Alzheimer's disease, amnesia, bipolar
CC disorder, dementia, depression, epilepsy, ischaemic cerebrovascular
CC disease, stroke, cerebral neoplasms, Pick's disease, Huntington's
CC disease and Parkinson's disease, demyelinating diseases, mental disorders
CC including mood, anxiety, Schizophrenia and seasonal affective disorder,
CC muscle disorder including cardiomyopathy, myocarditis, polymyositis,
CC dermatomyositis, arrhythmias and asthma and immunological disorders
CC including AIDS, adult respiratory distress syndrome (ARDS), allergies,
CC anaemia, diabetes mellitus, rheumatoid arthritis, scleroderma, Sjogren's
CC syndrome, systemic lupus erythematosus and other diseases including
CC sickle cell anaemia, Wilson's disease, cataracts, infertility, pulmonary
CC artery stenosis, Grave's disease, Cushing's disease, Addison's disease,
CC glucose-galactose malabsorption syndrome, hypercholesterolaemia, cancers
CC psoriasis and viral, bacterial, fungal, helminthic and protozoal
CC infections. TRICH DNA is useful in gene therapy and in diagnostic
CC purposes.
XX
SQ Sequence 3179 BP; 562 A; 1036 C; 962 G; 619 T; 0 other;

Query Match 12.4%; Score 32; DB 22; Length 3179;
Best Local Similarity 42.8%; Pred. No. 0.6;
Matches 65; Conservative 12; Mismatches 75; Indels 0; Gaps 0;

```
Qy      9 gaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaugacgucccu 68
          || ||: |:  |      |  ::  |||     |||: | |   | ||  : ||  |||
Db   2836 GAGCATTGTCCAAGCTGGCTCTTGGGGGGTCCCCCATTGGCCACAAGGCCTCACCCCCCA 2777

Qy     69 ucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcuguguggggaagca 128
            | |  |||  | ||||  |: |  :|    || ||: | ||| |:| | || |||
Db   2776 CCCCACCCCCGCACCAGAACCACCTTGATTGCAGCAGAAGTGCAGGCAGTGGGTGACGCA 2717

Qy    129 gucaguauaauucccgucgugugguggugacgc 160
          |   |   |  :|||     | | | :|  ||
Db   2716 GGGGGCTAAGCTCCCCCAAAGCGGGTTGGGGC 2685
```

RESULT 7
AAC81741/c
ID AAC81741 standard; DNA; 522 BP.
XX
AC AAC81741;
XX
DT 26-FEB-2001 (first entry)
XX
DE Human secreted protein coding sequence SEQ ID NO: 42.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ss.
XX
OS Homo sapiens.
XX
PN WO200061627-A1.
XX
PD 19-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US09067.
XX
PR 09-APR-1999; 99US-0128697.
PR 20-JAN-2000; 2000US-0176929.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-647419/62.
DR P-PSDB; AAB45417.
XX

PT New nucleic acid molecules encoding 49 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS Claim 1; Page 446; 521pp; English.
XX
CC The invention relates to the isolation of genes AAC81710-C81758 encoding
CC 49 human secreted proteins AAB45386-B54534. The genes can be used to
CC generate fusion proteins by linking to the gene for the human
CC immunoglobulin G Fc portion (SEQID1) for increasing the stability of
CC the fusion protein as compared to the human protein only. The genes and
CC proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
XX
SQ Sequence 522 BP; 128 A; 140 C; 135 G; 119 T; 0 other;

Query Match 11.7%; Score 30.4; DB 21; Length 522;
Best Local Similarity 42.1%; Pred. No. 1.1;
Matches 64; Conservative 12; Mismatches 76; Indels 0; Gaps 0;

```
Qy       9 gaccaugguggaucagaaccguuucgggugaagccaugqucugaaggggaugacgucccu 68
           || ||: |:  |      |  ::  |||     |||: | |   | ||  : ||  |||
Db     157 GAGCATTGTCCAAGCTGGCTCTTGGGGGGTCCCCCATTGGCCACAAGGCCTCACCCCCCA 98

Qy      69 ucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcuguguggggaagca 128
             |  | |   ||| | ||||  |: |  :|    || ||: | ||  |:| | || |||
Db      97 CCCCACCCCCGCACCAGAACCACCTTGATTGCAGCAGAAGTGCAGGGAGTGGGTGACGCA 38

Qy     129 gucaguauaauucccgucgugugguggugacgc 160
           |   |   |  :|||     | | | :|  ||
Db      37 GGGGGCTAAGCTCCCCCAAAGCGGGTTGGGGC 6
```

RESULT 8
AAA93110/c
ID AAA93110 standard; cDNA; 2342 BP.
XX
AC AAA93110;
XX
DT 12-JAN-2001 (first entry)
XX
DE Human secreted protein coding sequence SEQ ID NO: 19.
XX
KW Human; secreted protein; cytokine; cell proliferation;
KW nutritional supplement; immune modulation; autoimmune disorder;
KW haematopoiesis regulation; tissue growth; haemostasis; inflammation; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 2..736
FT /*tag= a
FT /product= "secreted protein"
FT sig_peptide 32..70
FT /*tag= b
FT mat_peptide 71..733
FT /*tag= c
XX
PN WO200049134-A1.
XX
PD 24-AUG-2000.
XX
PF 18-FEB-2000; 2000WO-US04340.
XX
PR 19-FEB-1999; 99US-0120680.
PR 23-APR-1999; 99US-0298733.
PR 17-AUG-1999; 99US-0149639.
PR 23-SEP-1999; 99US-0155686.
PR 01-OCT-1999; 99US-0157247.
PR 29-NOV-1999; 99US-0167822.
PR 29-NOV-1999; 99US-0167823.
PR 15-FEB-2000; 2000US-0298733.
XX
PA (ALPH-) ALPHAGENE INC.
XX
PI Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;
XX
DR WPI; 2000-549267/50.
DR P-PSDB; AAB23610.
XX
PT New secreted proteins and polynucleotides encoding them, which are
PT derived from Homo sapiens, useful for therapy, diagnosis, and research,
PT as well as nutritional sources or supplements -
XX
PS Claim 28; Page 253; 309pp; English.
XX
CC The present sequence is the coding sequence for a human secreted protein.
CC It was isolated from an adult pancreas cDNA library. The proteins
CC and coding sequences of the invention can be used in the isolation of
CC similar genes and proteins, in the elucidation of their function in vivo,
CC and to treat a number of conditions. It is possible that they may have
CC uses as nutritional supplements, as cytokine or cell proliferation
CC factors, in immune modulation, where they may be used to treat immune and
CC autoimmune diseases, as haematopoiesis regulators (treating myeloid or
CC lymphoid cell deficiencies), in the promotion of tissue growth, they may
CC have chemokine or chemotactic activity, haemostatic or thrombolytic
CC activity, or anti-inflammatory activity.
XX
SQ Sequence 2342 BP; 451 A; 742 C; 679 G; 470 T; 0 other;

Query Match 11.7%; Score 30.4; DB 21; Length 2342;
Best Local Similarity 42.1%; Pred. No. 1.9;
Matches 64; Conservative 12; Mismatches 76; Indels 0; Gaps 0;

```
Qy       9 gaccaugguggaucagaaccguuucgggugaagccaugqucugaaggggaugacgucccu 68
           || ||: |:  |      |  ::  |||     |||: | |   | ||  : ||  |||
Db    1980 GAGCATTGTCCAAGCTGGCTCTTGGGGGGTCCCCCATTGGCCACAAGGCCTCACCCCCCA 1921

Qy      69 ucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcuguguggggaagca 128
             |  | |   ||| | ||||  |: |  :|    || ||: | ||  |:| | || |||
Db    1920 CCCCACCCCCGCACCAGAACCACCTTGATTGCAGCAGAAGTGCAGGGAGTGGGTGACGCA 1861

Qy     129 gucaguauaauucccgucgugugguggugacgc 160
           |   |   |  :|||     | | | :|  ||
Db    1860 GGGGGCTAAGCTCCCCCAAAGCGGGTTGGGGC 1829
```

RESULT 9
AAZ23900
ID AAZ23900 standard; DNA; 49999 BP.
XX
AC AAZ23900;
XX
DT 25-JAN-2000 (first entry)
XX
DE Human LOBO homologue genomic DNA fragment 2.
XX
KW LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphysal dysplasia; achondroplasia; human; ds.
XX
OS Homo sapiens.

XX
PN WO9950284-A2.
XX
PD 07-OCT-1999.
XX
PF 26-MAR-1999; 99WO-EP02055.
XX
PR 27-MAR-1998; 98DE-1013799.
XX
PA (ROSE/) ROSENTHAL A.
XX
PI Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX
DR WPI; 1999-601320/51.
XX
PT Nucleic acids encoding proteins which influence bone development,
PT useful for treating and studying bone disorders -
XX
PS Example 3; Page 244-271; 391pp; German.
XX
CC This invention describes novel nucleic acids (I; designated LOBO (long
CC bones)) encoding proteins influencing bone development in mammals. The
CC proteins of the invention reduce and/or inactivate bone extension (i.e.
CC development), with exception of the skull and have osteopathic activity.
CC The nucleic acid molecules, proteins and antibodies can be used in
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC and nucleic acid molecules, etc. are useful for production of transgenic
CC animals, especially a transgenic mouse for the study of diseases
CC associated with bone development, e.g. spondyloepiphysal dysplasia and
CC achondroplasia. This sequence encodes a human LOBO protein described
CC in the method of the invention.
XX
SQ Sequence 49999 BP; 11432 A; 13017 C; 12505 G; 13045 T; 0 other;

Query Match 11.6%; Score 30; DB 20; Length 49999;
Best Local Similarity 39.8%; Pred. No. 7.9;
Matches 47; Conservative 16; Mismatches 55; Indels 0; Gaps 0;

Qy 14 ugguggaucagaaccguuucgggugaagccauggucugaaggggaugacgucccuucugg 73
:||:|||:|||| | :: | | :| | ||: |:| | || || || : : :|
Db 24459 tggtggatcagagcacttgcagttggggacattgtcagtagtagaagatctgagtgatgc 24518

Qy 74 cucauccacaaaaaccgucucgggugggugaggaguccuggcuguguggggaagcaguc 131
|:| :|| |: | | || || | || : |||| | ||:|
Db 24519 ctcttcctggctgctgcacttgaggaggggacagagcagggtgttcctggatgaagtc 24576

RESULT 10
AAS21771/c
ID AAS21771 standard; DNA; 24183 BP.
XX
AC AAS21771;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human gene for collagen COL9A1.
XX
KW Human; collagen; COL1A1; COL1A2; COL9A1; COL9A2; COL9A3; ds;
KW osteoporosis; multiple epiphyseal dysplasia; osteogenesis imperfecta;
KW shortness of stature; low bone density; gene therapy.
XX
OS Homo sapiens.
XX
PN US6265157-B1.
XX
PD 24-JUL-2001.
XX
PF 03-OCT-1997; 97US-0943731.
XX
PR 03-DEC-1991; 91US-0803628.
PR 13-MAR-1994; 94US-0212322.
XX
PA (UYAL-) UNIV ALLEGHENY HEALTH SCI.
PA (UYJE-) UNIV JEFFERSON THOMAS.
PA (UYOU-) UNIV OULU.
XX
PI Prockop DJ, Spotila LD, Deltas CD, Sereda L, Westerhausen Larson A;
PI Pack M, Colige A, Early J, Koerkkoe J, Ala-Kokko L, Annunen S;
PI Pihlajamaa T, Vuoristo M, Paassilta P;
XX
DR WPI; 2001-432201/46.
XX
PT Detecting collagen gene alteration, useful for diagnosing osteoporosis,
PT multiple epiphyseal dysplasia, osteogenesis imperfecta, shortness of
PT stature and low bone density in humans -
XX
PS Example 5; Fig 10A; 617pp; English.
XX
CC The invention relates to Detecting a collagen gene alteration associated
CC with a pathological condition in a human subject by obtaining from the
CC subject a sample nucleic acid containing a portion of at least 15
CC consecutive nucleotides of the segment of the COL1A1 gene extending in
CC the 5' to 3' direction from 78 nucleotides of intron 27 located adjacent
CC exon 28 through the 3' end of intron 51, where the portion contains an
CC intronic nucleotide and a first and second site, determining the sequence
CC of the portion and comparing the sequence of the portion with the
CC corresponding consensus sequence of the COL1A1 gene where a difference
CC between the sequence of the portion and the consensus sequence indicates
CC the presence of the collagen alteration in the subject. The method is
CC used for detecting abnormalities in a COL1 or COL9 gene is useful for
CC determining whether a subject is afflicted with pathological conditions
CC associated with an altered collagen gene such as osteoporosis, multiple
CC epiphyseal dysplasia, osteogenesis imperfecta, shortness of stature and
CC low bone density. Identification of an abnormality in a collagen gene is
CC also useful for designing a therapeutic nucleotide or gene therapy agent
CC which can be administered to the subject to correct or alleviate the
CC abnormality. The method is useful for detecting mutations in both the
CC coding and non-coding sequences of any of the COL1 or COL9 genes.
CC Therefore the method can be used to detect collagen gene alterations
CC which affect either the primary sequence of a collagen protein chain,
CC splicing of the mRNA encoding such chains or regulation of expression of
CC the genes encoding such chains. The present sequence is a collagen gene
CC of the invention.
XX
SQ Sequence 24183 BP; 6756 A; 4685 C; 5006 G; 7607 T; 129 other;

Query Match 11.5%; Score 29.8; DB 22; Length 24183;
Best Local Similarity 38.1%; Pred. No. 7.1;
Matches 43; Conservative 18; Mismatches 52; Indels 0; Gaps 0;

Qy 120 ugggaagcagucaguauaauucccgucgugugguggugacgccucacgacguauuuguccg 179
: | | | ||:||| | : |||: : | :| | |:|| | | ::: :|
Db 4413 TTGCAGGTAGTCAGGGGACTCACCGTTATTCTGGCTGGCAGCTCATGGCAAGTTTCTCTC 4354

Qy 180 cugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggag 232
|:| || | | |: ||| || : | | :: |::| || ||||
Db 4353 CTGGGCCGCAGGGGGTCACAATGGATCAGCATCCATTGAAGTTCAAACTGGAG 4301

RESULT 11
AAH20379/c
ID AAH20379 standard; cDNA; 1933 BP.
XX
AC AAH20379;
XX
DT 03-AUG-2001 (first entry)
XX
DE Human RNA-binding protein 33 cDNA.
XX
KW Human; RNA-binding protein 33; tumour; haemopathy; HIV infection;
KW immunological disease; inflammatory disease; ss.
XX
OS Homo sapiens.
XX

```
FH   Key             Location/Qualifiers
FT   CDS             853..1749
FT                   /*tag= a
FT                   /product= "RNA-binding protein 33"
XX
PN   WO200130839-A1.
XX
PD   03-MAY-2001.
XX
PF   27-OCT-2000; 2000WO-CN00381.
XX
PR   28-OCT-1999;   99CN-0119885.
XX
PA   (SHAN-) SHANGHAI BIO ROAD GENE DEV LTD.
XX
PI   Mao Y,  Xie Y;
XX
DR   WPI; 2001-291030/30.
DR   P-PSDB; AAB81198.
XX
PT   Human RNA-binding protein 33 and its encoding polynucleotides are
PT   applicable in diagnosis and treatment of malignant tumors, hemopathy,
PT   HIV infection, immunological diseases and various inflammations -
XX
PS   Claim 6; Page 24-25; 32pp; Chinese.
XX
CC   This invention relates to human RNA-binding protein 33 and the cDNA
CC   encoding it. The invention includes a vector containing the cDNA, host
CC   cells transformed or transduced with the vector, and an antibody that can
CC   specifically bind with the polypeptide. The polypeptide and
CC   polynucleotide it is encoded by are applicable in diagnosis and treatment
CC   of malignant tumours, haemopathy, HIV infection, immunological diseases
CC   and various inflammatory diseases and disorders. The present sequence
CC   represents cDNA encoding human RNA-binding protein 33.
XX
SQ   Sequence 1933 BP; 632 A; 397 C; 398 G; 506 T; 0 other;
```

```
  Query Match             11.4%;  Score 29.4;  DB 22;  Length 1933;
  Best Local Similarity   38.9%;  Pred. No. 3.9;
  Matches   79;  Conservative   23;  Mismatches   96;  Indels    5;  Gaps    1;

Qy      35 ggugaagccauggucugaaggggaugacgucccuucuggcucauccacaaaaaccgucuc 94
           || | |  ||: |  :| |  |  : |     |::| ||| |       ||  |  |   |
Db    1596 GGGGTACTCATAGGGTGTATAGATTAAACCAGCTTCGGGCCCTGGAGGAACTATTGCAGC 1537

Qy      95 gggugggugaggaguccuggcuguguggggaagcagucaguauaauuucccguc-----gug 149
           | :|||:|||||: | |  :|    |  ||| |:| |:|:   :|   :|
Db    1536 AGTTGGGTGAGGAGTTCCGTTTGGCATGACAGCGGTCTGTATTTGTCTGATCAAAGGCAT 1477

Qy     150 uguggugacgccucacgacguauuuguccgcugugcagagcguaguaccaagggcugcac 209
           : :||:   |||    | ||:|   |: ||| | ||||   |: | |  |  |||
Db    1476 TATGGTAGGGCCAGCTGGCGTAGGAGTACGCAGGGCAGCTGGTGGGAGAACCGGCGCAGG 1417

Qy     210 cccgguuuuuguuccaagcggag 232
           ||| |:  :  : |   | | ||
Db    1416 CCCAGTAATGATCCTTGGAGCAG 1394
```

```
RESULT  12
AAH15309/c
ID   AAH15309 standard; cDNA; 1993 BP.
XX
AC   AAH15309;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:13465.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
```

```
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 13465; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 1993 BP; 632 A; 419 C; 419 G; 523 T; 0 other;
```

```
  Query Match             11.4%;  Score 29.4;  DB 22;  Length 1993;
  Best Local Similarity   38.9%;  Pred. No. 4;
  Matches   79;  Conservative   23;  Mismatches   96;  Indels    5;  Gaps    1;

Qy      35 ggugaagccauggucugaaggggaugacgucccuucuggcucauccacaaaaaccgucuc 94
           || | |  ||: |  :| |  |  : |     |::| ||| |       ||  |  |   |
Db     831 GGGGTACTCATAGGGTGTATAGATTAAACCAGCTTCGGGCCCTGGAGGAACTATTGCAGC 772

Qy      95 gggugggugaggaguccuggcuguguggggaagcagucaguauaauuucccguc-----gug 149
           | :|||:|||||: | |  :|    |  ||| |:| |:|:   :|   :|
Db     771 AGTTGGGTGAGGAGTTCCGTTTGGCATGACAGCGGTCTGTATTTGTCTGATCAAAGGCAT 712

Qy     150 uguggugacgccucacgacguauuuguccgcugugcagagcguaguaccaagggcugcac 209
           : :||:   |||    | ||:|   |: ||| | ||||   |: | |  |  |||
Db     711 TATGGTAGGGCCAGCTGGCGTAGGAGTACGCAGGGCAGCTGGTGGGAGAACCGGCGCAGG 652

Qy     210 cccgguuuuuguuccaagcggag 232
           ||| |:  :  : |   | | ||
Db     651 CCCAGTAATGATCCTTGGAGCAG 629
```

```
RESULT  13
AAA49680/c
ID   AAA49680 standard; DNA; 2092 BP.
XX
AC   AAA49680;
XX
DT   25-SEP-2000  (first entry)
XX
DE   Human quaking splice variant 5 (Qk5).
XX
KW   Quaking splice variant 5; Qk5 gene; human;
KW   differential gene expression; movement disorder; motor pathology;
KW   dystonia; dyskinesia; therapy; diagnosis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             28..1053
FT                   /*tag=  a
XX
PN   WO200037685-A2.
XX
PD   29-JUN-2000.
XX
PF   21-DEC-1999;   99WO-US30727.
XX
PR   21-DEC-1998;   98US-0113127.
PR   20-DEC-1999;   99US-0468253.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Gould-Rothberg B;
XX
DR   WPI; 2000-442696/38.
DR   P-PSDB; AAY95336.
XX
PT   Identifying pyschotropic agents useful in the treatment of e.g.
PT   schizophrenia comprises comparing the expression patterns of HALO genes
PT   in test and control cell populations -
XX
PS   Disclosure; Fig 1; 93pp; English.
XX
CC   The present sequence of the human quaking splice variant 5 (Qk5),
CC   a homologue of the mouse quaking type 1 gene, which is a member of
CC   the STAR (signal transduction and activator of RNA) class of
CC   proteins.  Human Qk5 was isolated following the identification of a
CC   rat homologue (see AAA49679) of mouse quaking.  The rat gene was
CC   shown to be differentially expressed in response to the neuroleptic
CC   haloperidol.  The human Qk5 gene is useful as a marker for the
CC   onset of tardive dyskinesia/dystonias in human subjects taking
CC   neuroleptics.  Determination of its level of expression, and
CC   manipulation of its level of expression, are used in methods of
CC   diagnosing or determining susceptibility to a movement disorder,
CC   and in methods of preventing or delaying the onset of a motor
CC   pathology, especially dystonia.
XX
SQ   Sequence 2092 BP; 683 A; 433 C; 449 G; 527 T; 0 other;


  Query Match             11.4%;  Score 29.4;  DB 21;  Length 2092;
  Best Local Similarity   38.9%;  Pred. No. 4;
  Matches   79;  Conservative   23;  Mismatches   96;  Indels    5;  Gaps    1;

Qy      35 ggugaagccauggucugaaggggaugacgucccuucuggcucauccacaaaaaccgucuc 94
           || | |  ||: |  :| |  |  : |     |::| ||| |      ||  |  |   |
Db     900 GGGGTACTCATAGGGTGTATAGATTAAACCAGCTTCGGGCCCTGGAGGAACTATTGCAGC 841

Qy      95 gggugggugaggaguccuggcuguguggggaagcagucaguauaauuccccguc-----gug 149
             | :|||:|||||| : |  :|    |  ||| |:| |:|:   :|   :|
Db     840 AGTTGGGTGAGGAGTTCCGTTTGGCATGACAGCGGTCTGTATTTGTCTGATCAAAGGCAT 781

Qy     150 uguggugacgccucacgacguauuuguccgcugugcagagcguaguaccaagggcugcac 209
           : :||:   |||    | ||:|   |: ||| | ||||   |: | |  |  |||
```

```
Db     780 TATGGTAGGGCCAGCTGGCGTAGGAGTACGCAGGGCAGCTGGTGGGAGAACCGGCGCAGG 721

Qy     210 cccgguuuuuguuccaagcggag 232
           ||| |:  :  : |   | | ||
Db     720 CCCAGTAATGATCCTTGGAGCAG 698


RESULT  14
AAA49681/c
ID   AAA49681 standard; DNA; 2472 BP.
XX
AC   AAA49681;
XX
DT   25-SEP-2000  (first entry)
XX
DE   Human quaking splice variant 7 (Qk7).
XX
KW   Quaking splice variant 7; Qk7 gene; human;
KW   differential gene expression; movement disorder; motor pathology;
KW   dystonia; dyskinesia; therapy; diagnosis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             6..983
FT                   /*tag=  a
XX
PN   WO200037685-A2.
XX
PD   29-JUN-2000.
XX
PF   21-DEC-1999;   99WO-US30727.
XX
PR   21-DEC-1998;   98US-0113127.
PR   20-DEC-1999;   99US-0468253.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Gould-Rothberg B;
XX
DR   WPI; 2000-442696/38.
DR   P-PSDB; AAY95335.
XX
PT   Identifying pyschotropic agents useful in the treatment of e.g.
PT   schizophrenia comprises comparing the expression patterns of HALO genes
PT   in test and control cell populations -
XX
PS   Disclosure; Fig 2; 93pp; English.
XX
CC   The present sequence of the human quaking splice variant 7 (Qk7),
CC   a homologue of the mouse quaking type 1 gene, which is a member of
CC   the STAR (signal transduction and activator of RNA) class of
CC   proteins.  Human Qk7 was isolated following the identification of a
CC   rat homologue (see AAA49679) of mouse quaking.  The rat gene was
CC   shown to be differentially expressed in response to the neuroleptic
CC   haloperidol.  The human Qk5 gene is useful as a marker for the
CC   onset of tardive dyskinesia/dystonias in human subjects taking
CC   neuroleptics.  Determination of its level of expression, and
CC   manipulation of its level of expression, are used in methods of
CC   diagnosing or determining susceptibility to a movement disorder,
CC   and in methods of preventing or delaying the onset of a motor
CC   pathology, especially dystonia.
XX
SQ   Sequence 2472 BP; 754 A; 490 C; 500 G; 728 T; 0 other;


  Query Match             11.4%;  Score 29.4;  DB 21;  Length 2472;
  Best Local Similarity   38.9%;  Pred. No. 4.3;
  Matches   79;  Conservative   23;  Mismatches   96;  Indels    5;  Gaps    1;

Qy      35 ggugaagccauggucugaaggggaugacgucccuucuggcucauccacaaaaaccgucuc 94
           || | |  ||: |  :| |  |  : |     |::| ||| |      ||  |  |   |
Db     878 GGGGTACTCATAGGGTGTATAGATTAAACCAGCTTCGGGCCCTGGAGGAACTATTGCAGC 819
```

```
Qy      95 gggugggugaggaguccuggcuguguugggaagcagucaguauaauucccguc-----gug 149
           | :|||:||||||: | |  :|    |  ||| ||:| |:|:   :|   :|
Db     818 AGTTGGGTGAGGAGTTCCGTTTGGCATGACAGCGGTCTGTATTTGTCTGATCAAAGGCAT 759

Qy     150 uguggugacgccucacgacguauuuguccgcugugcagagcguaguaccaagggcugcac 209
           : :||:   |||    | ||:|   |: ||| | |||||   |: | |  |  |||
Db     758 TATGGTAGGGCCAGCTGGCGTAGGAGTACGCAGGGCAGCTGGTGGGAGAACCGGCGCAGG 699

Qy     210 cccgguuuuuguuccaagcggag 232
           ||| |:  :  : |   | | ||
Db     698 CCCAGTAATGATCCTTGGAGCAG 676


RESULT  15
AAC39642
ID   AAC39642 standard; DNA; 775 BP.
XX
AC   AAC39642;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 25380.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
```

PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.2%; Score 29; DB 21; Length 775;
Best Local Similarity 45.9%; Pred. No. 3.9;

Matches 39; Conservative 11; Mismatches 35; Indels 0; Gaps 0;

```
Qy      47 gucugaaggggaugacgucccuucuggcucauccacaaaaaccgucucgggugggugagg 106
           |:   || |   :|  |:|    |:|| : |:|||||  |||  :|   ||:|| :|| |
Db     124 gttccaaagctgtggagtcggcgctggatgatccacaggaacgatcggaggtggttgaag 183

Qy     107 aguccuggcugugugggaagcaguc 131
           |  || |||  :|  |    |||:|
Db     184 aacccaggcgctggagatctcagtc 208
```

Search completed: February 16, 2002, 16:15:33
Job time: 6441 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: February 16, 2002, 15:14:37 ; Search time 84.06 Seconds
(without alignments)
697.809 Million cell updates/sec

Title: US-09-587-653-1
Perfect score: 259
Sequence: 1 aguuuggcgaccauggugga.....cccgcuuggaauuaaaaacu 259

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	29.8	11.5	24183	4	US-08-943-731-3	Sequence 3, Appli
	2	29.2	11.3	1368	1	US-07-677-413-1	Sequence 1, Appli
	3	29.2	11.3	1571	1	US-07-677-413-2	Sequence 2, Appli
	4	28	10.8	1723	1	US-07-841-646-28	Sequence 28, Appl
	5	28	10.8	1723	1	US-07-901-703-10	Sequence 10, Appl
	6	28	10.8	1723	1	US-08-147-023-28	Sequence 28, Appl
	7	28	10.8	1723	1	US-08-206-864-3	Sequence 3, Appli
	8	28	10.8	1723	1	US-08-278-729A-20	Sequence 20, Appl
	9	28	10.8	1723	1	US-08-480-528A-7	Sequence 7, Appli
	10	28	10.8	1723	1	US-08-479-666-7	Sequence 7, Appli
	11	28	10.8	1723	1	US-08-155-343A-20	Sequence 20, Appl
	12	28	10.8	1723	1	US-08-406-672-20	Sequence 20, Appl
	13	28	10.8	1723	1	US-08-643-563A-20	Sequence 20, Appl
	14	28	10.8	1723	1	US-08-447-570-28	Sequence 28, Appl
	15	28	10.8	1723	1	US-08-643-763A-20	Sequence 20, Appl
	16	28	10.8	1723	1	US-08-462-623-20	Sequence 20, Appl
	17	28	10.8	1723	1	US-08-451-953A-20	Sequence 20, Appl
	18	28	10.8	1723	2	US-08-459-346-5	Sequence 5, Appli
	19	28	10.8	1723	2	US-08-445-468A-20	Sequence 20, Appl
	20	28	10.8	1723	2	US-08-901-200A-7	Sequence 7, Appli
	21	28	10.8	1723	2	US-08-449-700-28	Sequence 28, Appl
	22	28	10.8	1723	2	US-08-449-699A-28	Sequence 28, Appl
	23	28	10.8	1723	2	US-08-461-397A-20	Sequence 20, Appl
	24	28	10.8	1723	2	US-08-912-088-20	Sequence 20, Appl
	25	28	10.8	1723	3	US-08-278-730A-20	Sequence 20, Appl
	26	28	10.8	1723	3	US-08-889-419-5	Sequence 5, Appli
	27	28	10.8	1723	3	US-08-445-467-20	Sequence 20, Appl
	28	28	10.8	1723	3	US-08-480-515A-20	Sequence 20, Appl
	29	28	10.8	1723	3	US-09-219-391-7	Sequence 7, Appli
	30	28	10.8	1723	5	PCT-US92-01968-20	Sequence 20, Appl
	31	28	10.8	1723	5	PCT-US93-05446-10	Sequence 10, Appl
	32	28	10.8	1723	5	PCT-US93-07189-5	Sequence 5, Appli
	33	28	10.8	1723	5	PCT-US93-07190-20	Sequence 20, Appl
	34	28	10.8	1723	5	PCT-US93-07231-20	Sequence 20, Appl
	35	28	10.8	1723	5	PCT-US93-08742-20	Sequence 20, Appl
	36	28	10.8	1723	5	PCT-US93-08808-20	Sequence 20, Appl
	37	28	10.8	1723	5	PCT-US93-08885-20	Sequence 20, Appl
	38	28	10.8	1723	5	PCT-US93-10520-7	Sequence 7, Appli
	39	28	10.8	1941	5	PCT-US91-07635-5	Sequence 5, Appli
c	40	28	10.8	10807	1	US-08-206-176-7	Sequence 7, Appli
c	41	28	10.8	10807	2	US-08-756-506-5	Sequence 5, Appli
	42	27.8	10.7	492	3	US-09-354-129-11	Sequence 11, Appl
	43	27	10.4	377	4	US-09-112-584-2	Sequence 2, Appli
c	44	26.4	10.2	31571	1	US-08-323-443B-1	Sequence 1, Appli
	45	26.2	10.1	1917	4	US-08-808-346-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT   1
US-08-943-731-3/c
; Sequence 3, Application US/08943731
; Patent No. 6265157
;  GENERAL INFORMATION:
;    APPLICANT:  PROCKOP, DARWIN J.
;    APPLICANT:  SPOTILA, LORETTA D.
;    APPLICANT:  DELTAS, CONSTANTINOS D.
;    APPLICANT:  SEREDA, LARISA
;    APPLICANT:  LARSON, ANDREA W.
;    APPLICANT:  PACK, MICHAEL
;    APPLICANT:  COLIGE, ALAIN
;    APPLICANT:  EARLY, JAMES
;    APPLICANT:  KORKKO, JARMO
;    APPLICANT:  ALA-KOKKO, LEENA, et al.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR DETECTING
;    TITLE OF INVENTION:  ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
;    NUMBER OF SEQUENCES:  666
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PANITCH SCHWARZE JACOBS & NADEL, P.C.
;      STREET:  ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
;      STREET:  FLR.
;      CITY:  PHILADELPHIA
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-7086
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/943,731
;      FILING DATE:  03-OCT-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/212,322
;      FILING DATE:  14-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/803,628
;      FILING DATE:  03-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DOYLE LEARY Ph.D., KATHRYN
;      REGISTRATION NUMBER:  36,317
;      REFERENCE/DOCKET NUMBER:  9598-27
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-965-1284
;      TELEFAX:  215-567-2991
;      TELEX:  831-494
;  INFORMATION FOR SEQ ID NO:  3:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  24183 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-943-731-3


  Query Match             11.5%;  Score 29.8;  DB 4;  Length 24183;
  Best Local Similarity   38.1%;  Pred. No. 1.5;
  Matches   43;  Conservative   18;  Mismatches   52;  Indels    0;  Gaps    0;

Qy     120 ugggaagcagucaguauaaauucccgucgugugguggugacgccucacgacguauuuguccg 179
           : | | | ||:|||   | :  |||:  :    | :| |  |:|| | |   ::: :|
Db    4413 TTGCAGGTAGTCAGGGGACTCACCGTTATTCTGGCTGGCAGCTCATGGCAAGTTTCTCTC 4354

Qy     180 cugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggag 232
           |:| || |   |  |:  ||| || :   |  | ::   |::| ||  ||||
Db    4353 CTGGGCCGCAGGGGGTCACAATGGATCAGCATCCATTGAAGTTCAAACTGGAG 4301


RESULT   2
US-07-677-413-1
; Sequence 1, Application US/07677413
; Patent No. 5260198
;  GENERAL INFORMATION:
;    APPLICANT:  Iwamori, Satoru
;    APPLICANT:  Oikawa, Toshihiro
;    APPLICANT:  Yoshino, Setsuo
;    APPLICANT:  Ishiwata, Ken-ichi
;    TITLE OF INVENTION:  Cloned Tyrosine Phenol-lyase Gene,
;    TITLE OF INVENTION:  Recombinant Plasmid Containing the Same and Escherichia
;    TITLE OF INVENTION:  Coli Transformed with the Same
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Teresa Stanek Rea
;      STREET:  699 Prince Street
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/677,413
;      FILING DATE:  19910429
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rea, Teresa S
;      REGISTRATION NUMBER:  30,427
;      REFERENCE/DOCKET NUMBER:  027066-005
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6620
;      TELEFAX:  703-836-2021
;      TELEX:  ITT 440 580 BDSM ALE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1368 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-07-677-413-1


  Query Match             11.3%;  Score 29.2;  DB 1;  Length 1368;
  Best Local Similarity   48.5%;  Pred. No. 0.77;
  Matches   32;  Conservative   11;  Mismatches   23;  Indels    0;  Gaps    0;

Qy     136 uaauucccgucgugugguggugacgccucacgacguauuuguccgcugugcagagcguagu 195
           : |: | ||:||: : :|   || ||: |||  | |:  |: | |:| |||||| : |
Db    1041 TGATGCTCGTCGTTTCTGTCCACACCTGACGCAGGATCAGTTCCCTGCGCAGAGCCTGGC 1100

Qy     196 accaag 201
           | | ||
Db    1101 AGCCAG 1106


RESULT   3
US-07-677-413-2
; Sequence 2, Application US/07677413
; Patent No. 5260198
;  GENERAL INFORMATION:
;    APPLICANT:  Iwamori, Satoru
;    APPLICANT:  Oikawa, Toshihiro
;    APPLICANT:  Yoshino, Setsuo
;    APPLICANT:  Ishiwata, Ken-ichi
;    TITLE OF INVENTION:  Cloned Tyrosine Phenol-lyase Gene,
;    TITLE OF INVENTION:  Recombinant Plasmid Containing the Same and Escherichia
;    TITLE OF INVENTION:  Coli Transformed with the Same
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Teresa Stanek Rea
;      STREET:  699 Prince Street
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/677,413
;      FILING DATE:  19910429
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rea, Teresa S
;      REGISTRATION NUMBER:  30,427
;      REFERENCE/DOCKET NUMBER:  027066-005
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6620
;      TELEFAX:  703-836-2021
;      TELEX:  ITT 440 580 BDSM ALE
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1571 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  176..1543
US-07-677-413-2


  Query Match             11.3%;  Score 29.2;  DB 1;  Length 1571;
  Best Local Similarity   48.5%;  Pred. No. 0.81;
  Matches   32;  Conservative   11;  Mismatches   23;  Indels    0;  Gaps    0;

Qy     136 uaauucccgucgugugguggugacgccucacgacguauuuguccgcugugcagagcguagu 195
           : |: | ||:||: : :|   || ||: |||  | |:  |: | |:| |||||| : |
Db    1216 TGATGCTCGTCGTTTCTGTCCACACCTGACGCAGGATCAGTTCCCTGCGCAGAGCCTGGC 1275

Qy     196 accaag 201
           | | ||
Db    1276 AGCCAG 1281


RESULT   4
```

```
US-07-841-646-28
; Sequence 28, Application US/07841646
; Patent No. 5266683
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  OSTEOGENIC DEVICES
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  U.S.A.
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/841,646
;      FILING DATE:  19920221
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 810,560
;      FILING DATE:  20-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 827,052
;      FILING DATE:  28-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 660,162
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,988
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,849
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 616,374
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 600,024
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 599,543
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 579,865
;      FILING DATE:  07-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 569,920
;      FILING DATE:  20-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 483,913
;      FILING DATE:  22-FEB-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 422,613
;      FILING DATE:  17-OCT-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 315,342
;      FILING DATE:  23-FEB-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 232,630
;      FILING DATE:  15-AUG-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 179,460
;      FILING DATE:  08-APR-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER, EDMUND R.
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-001CP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1723 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  490..1696
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "hOP2-PP"
;      OTHER INFORMATION:  /note= "hOP2 (cDNA)"
US-07-841-646-28


  Query Match             10.8%;  Score 28;  DB 1;  Length 1723;
  Best Local Similarity   47.0%;  Pred. No. 2.2;
  Matches   47;  Conservative    8;  Mismatches   45;  Indels    0;  Gaps    0;

Qy       1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccaugucugaaggggaug 60
           || :: |  | ||:||:||| |   ||||:  |  | |   ||| |  |   |  ||| |
Db     808 AGCTTCGTTAACATGGTGGAGCGAGACCGTGCCCTGGGCCACCAGGAGCCCCATTGGAAG 867

Qy      61 acgucccuucuggcucauccacaaaaaccgucucgggugg 100
             |: ||   : |  |   | || |  ||| |: ||| ||
Db     868 GAGTTCCGCTTTGACCTGACCCAGATCCCGGCTGGGGAGG 907


RESULT   5
US-07-901-703-10
; Sequence 10, Application US/07901703
; Patent No. 5344654
;  GENERAL INFORMATION:
;    APPLICANT:  RUEGER, DAVID C
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZAKAYNAK, ENGIN
;    TITLE OF INVENTION:  PROSTHETIC DEVICES HAVING ENHANCED
;    TITLE OF INVENTION:  OSTEOGENIC PROPERTIES
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  EXCHANGE PLACE, 53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/901,703
;      FILING DATE:  19920616
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER  ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  STK-057
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;  INFORMATION FOR SEQ ID NO:  10:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1723 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  490..1696
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "hOP2-PP"
;      OTHER INFORMATION:  /note= "hOP2 (cDNA)"
US-07-901-703-10


  Query Match             10.8%;  Score 28;  DB 1;  Length 1723;
  Best Local Similarity   47.0%;  Pred. No. 2.2;
  Matches   47;  Conservative    8;  Mismatches   45;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccaugucugaaggggaug 60
          || :: |  | ||:||:||| |   ||||:  |  | |   ||| |  |   |  ||| |
Db    808 AGCTTCGTTAACATGGTGGAGCGAGACCGTGCCCTGGGCCACCAGGAGCCCCATTGGAAG 867

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugg 100
            |: ||   : |  |   | || |  ||| |: ||| ||
Db    868 GAGTTCCGCTTTGACCTGACCCAGATCCCGGCTGGGGAGG 907


RESULT   6
US-08-147-023-28
; Sequence 28, Application US/08147023
; Patent No. 5468845
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  OSTEOGENIC DEVICES
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  U.S.A.
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/147,023
;      FILING DATE:  21-FEB-1992
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 810,560
;      FILING DATE:  20-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 827,052
;      FILING DATE:  28-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 660,162
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,988
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,849
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 616,374
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 600,024
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 599,543
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 579,865
;      FILING DATE:  07-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 569,920
;      FILING DATE:  20-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 483,913
;      FILING DATE:  22-FEB-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 422,613
;      FILING DATE:  17-OCT-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 315,342
;      FILING DATE:  23-FEB-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 232,630
;      FILING DATE:  15-AUG-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 179,460
;      FILING DATE:  08-APR-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER, EDMUND R.
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-001CP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1723 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  490..1696
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "hOP2-PP"
;      OTHER INFORMATION:  /note= "hOP2 (cDNA)"
US-08-147-023-28


  Query Match             10.8%;  Score 28;  DB 1;  Length 1723;
  Best Local Similarity   47.0%;  Pred. No. 2.2;
  Matches   47;  Conservative    8;  Mismatches   45;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccaugucugaaggggaug 60
          || :: |  | ||:||:||| |   ||||:  |  | |   ||| |  |   |  ||| |
Db    808 AGCTTCGTTAACATGGTGGAGCGAGACCGTGCCCTGGGCCACCAGGAGCCCCATTGGAAG 867

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugg 100
            |: ||   : |  |   | || |  ||| |: ||| ||
Db    868 GAGTTCCGCTTTGACCTGACCCAGATCCCGGCTGGGGAGG 907


RESULT   7
US-08-206-864-3
; Sequence 3, Application US/08206864
```

```
; Patent No. 5610021
;  GENERAL INFORMATION:
;    APPLICANT:  RUEGER, DAVID C
;    APPLICANT:  JONES, WILLIAM K
;    APPLICANT:  TUCKER, RONALD F
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR
;    TITLE OF INVENTION:  RECOMBINANT OSTEOGENIC PROTEIN PRODUCTION
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR/CREATIVE BIOMOLECULES,
;      ADDRESSEE:  INC.
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/206,864
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/027,070
;      FILING DATE:  04-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/841,646
;      FILING DATE:  21-FEB-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-096
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1723 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  490..1696
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "hOP2-PP"
;      OTHER INFORMATION:  /note= "hOP2 (cDNA)"
US-08-206-864-3


  Query Match             10.8%;  Score 28;  DB 1;  Length 1723;
  Best Local Similarity   47.0%;  Pred. No. 2.2;
  Matches   47;  Conservative    8;  Mismatches   45;  Indels    0;  Gaps    0;

Qy       1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
           || ::  |  | ||:||:||| |   ||||:  |  | |   ||| |  |   | ||| |
Db     808 AGCTTCGTTAACATGGTGGAGCGAGACCGTGCCCTGGGCCACCAGGAGCCCCATTGGAAG 867

Qy      61 acgucccuucuggcucauccacaaaaaccgucucgggugg 100
             |: ||   : |  |   | || |  ||| |: ||| ||
Db     868 GAGTTCCGCTTTGACCTGACCCAGATCCCGGCTGGGGAGG 907
```

```
RESULT   8
US-08-278-729A-20
; Sequence 20, Application US/08278729A
; Patent No. 5650276
;  GENERAL INFORMATION:
;    APPLICANT:  SMART, JOHN
;    APPLICANT:  OPPERMANN, HERMAN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    APPLICANT:  COHEN, CHARLES M.
;    TITLE OF INVENTION:  MORPHOGENIC PROTEIN SCREENING METHOD
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/278,729A
;      FILING DATE:  20-JUL-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER Esq., EDMUND R.
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-058CPFW
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (508) 435-9001
;      TELEFAX:  (508) 435-6951
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1723 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  490..1695
US-08-278-729A-20


  Query Match             10.8%;  Score 28;  DB 1;  Length 1723;
  Best Local Similarity   47.0%;  Pred. No. 2.2;
  Matches   47;  Conservative    8;  Mismatches   45;  Indels    0;  Gaps    0;

Qy       1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
           || ::  |  | ||:||:||| |   ||||:  |  | |   ||| |  |   | ||| |
Db     808 AGCTTCGTTAACATGGTGGAGCGAGACCGTGCCCTGGGCCACCAGGAGCCCCATTGGAAG 867

Qy      61 acgucccuucuggcucauccacaaaaaccgucucgggugg 100
             |: ||   : |  |   | || |  ||| |: ||| ||
Db     868 GAGTTCCGCTTTGACCTGACCCAGATCCCGGCTGGGGAGG 907
```

```
RESULT   9
US-08-480-528A-7
; Sequence 7, Application US/08480528A
; Patent No. 5652118
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
```

```
;    APPLICANT:  COHEN, CHARLES M.
;    TITLE OF INVENTION:  OP3-INDUCED MORPHOGENESIS
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,528A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FENTON Esq., GILLIAN M.
;      REGISTRATION NUMBER:  36,508
;      REFERENCE/DOCKET NUMBER:  CRP-076FW
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7560
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1723 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  490..1696
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "hOP2-PP"
;      OTHER INFORMATION:  /note= "hOP2 cDNA"
US-08-480-528A-7


  Query Match             10.8%;  Score 28;  DB 1;  Length 1723;
  Best Local Similarity   47.0%;  Pred. No. 2.2;
  Matches   47;  Conservative    8;  Mismatches   45;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccaugucugaaggggaug 60
          || :: |  | ||:||:||| |   ||||:  |  | |   ||| |  |   |  ||| |
Db    808 AGCTTCGTTAACATGGTGGAGCGAGACCGTGCCCTGGGCCACCAGGAGCCCCATTGGAAG 867

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugg 100
            |: ||   : |  |   | || |  ||| |: ||| ||
Db    868 GAGTTCCGCTTTGACCTGACCCAGATCCCGGCTGGGGAGG 907


RESULT  10
US-08-479-666-7
; Sequence 7, Application US/08479666
; Patent No. 5652337
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    APPLICANT:  COHEN, CHARLES M.
;    TITLE OF INVENTION:  OP3-INDUCED MORPHOGENESIS
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,666
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FENTON Esq., GILLIAN M.
;      REGISTRATION NUMBER:  36,508
;      REFERENCE/DOCKET NUMBER:  CRP-076DV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (508) 435-9001
;      TELEFAX:  (508) 435-6951
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1723 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  490..1696
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "hOP2-PP"
;      OTHER INFORMATION:  /note= "hOP2 cDNA"
US-08-479-666-7


  Query Match             10.8%;  Score 28;  DB 1;  Length 1723;
  Best Local Similarity   47.0%;  Pred. No. 2.2;
  Matches   47;  Conservative    8;  Mismatches   45;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccaugucugaaggggaug 60
          || :: |  | ||:||:||| |   ||||:  |  | |   ||| |  |   |  ||| |
Db    808 AGCTTCGTTAACATGGTGGAGCGAGACCGTGCCCTGGGCCACCAGGAGCCCCATTGGAAG 867

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugg 100
            |: ||   : |  |   | || |  ||| |: ||| ||
Db    868 GAGTTCCGCTTTGACCTGACCCAGATCCCGGCTGGGGAGG 907


RESULT  11
US-08-155-343A-20
; Sequence 20, Application US/08155343A
; Patent No. 5656593
;  GENERAL INFORMATION:
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  OPPERMANN, HERMAN
;    APPLICANT:  COHEN, CHARLES M.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  MORPHOGENIC-INDUCED PERIODONTAL TISSUE
;    TITLE OF INVENTION:  REGENERATION.
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US/08/155,343A
;       FILING DATE:  15-NOV-1993
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  FENTON Esq., GILLIAN M.
;       REGISTRATION NUMBER:  36,508
;       REFERENCE/DOCKET NUMBER:  CRP-067FW
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 248-7560
;       TELEFAX:  (617) 248-7100
;   INFORMATION FOR SEQ ID NO:  20:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1723 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  490..1695
US-08-155-343A-20


  Query Match             10.8%;  Score 28;  DB 1;  Length 1723;
  Best Local Similarity   47.0%;  Pred. No. 2.2;
  Matches   47;  Conservative    8;  Mismatches   45;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccaugqucugaaggggaug 60
          || :: |  | ||:||:||| |   ||||:  | | |   ||| |  |   |  ||| |
Db    808 AGCTTCGTTAACATGGTGGAGCGAGACCGTGCCCTGGGCCACCAGGAGCCCCATTGGAAG 867

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugg 100
            |: ||   : |  |   | || |  ||| |:|||| ||
Db    868 GAGTTCCGCTTTGACCTGACCCAGATCCCGGCTGGGGAGG 907


RESULT  12
US-08-406-672-20
; Sequence 20, Application US/08406672
; Patent No. 56744844
;  GENERAL INFORMATION:
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  COHEN, CHARLES M.
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  TREATMENT TO PREVENT LOSS OF AND/OR
;    TITLE OF INVENTION:  INCREASE BONE MASS IN METABOLIC BONE DISEASES
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/406,672
;      FILING DATE:  20-MAR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 752,857
;      FILING DATE:  30-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 667,274
;      FILING DATE:  11-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FENTON Esq., GILLIAN M.
;      REGISTRATION NUMBER:  36,508
;      REFERENCE/DOCKET NUMBER:  CRP-060CN
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7560
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1723 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  490..1695
US-08-406-672-20


  Query Match             10.8%;  Score 28;  DB 1;  Length 1723;
  Best Local Similarity   47.0%;  Pred. No. 2.2;
  Matches   47;  Conservative    8;  Mismatches   45;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccaugqucugaaggggaug 60
          || :: |  | ||:||:||| |   ||||:  | | |   ||| |  |   |  ||| |
Db    808 AGCTTCGTTAACATGGTGGAGCGAGACCGTGCCCTGGGCCACCAGGAGCCCCATTGGAAG 867

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugg 100
            |: ||   : |  |   | || |  ||| |:|||| ||
Db    868 GAGTTCCGCTTTGACCTGACCCAGATCCCGGCTGGGGAGG 907


RESULT  13
US-08-643-563A-20
; Sequence 20, Application US/08643563A
; Patent No. 5707810
;  GENERAL INFORMATION:
;    APPLICANT:  SMART, JOHN
;    APPLICANT:  OPPERMANN, HERMAN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    APPLICANT:  COHEN, CHARLES M.
;    TITLE OF INVENTION:  MORPHOGENIC PROTEIN SCREENING METHOD
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/643,563A
;      FILING DATE:  06-MAY-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  TWOMEY Esq., MICHAEL J.
;      REGISTRATION NUMBER:  38,349
;      REFERENCE/DOCKET NUMBER:  CRP-058CN2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (508) 435-9001
;      TELEFAX:  (508) 435-6951
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1723 base pairs
;      TYPE:  nucleic acid
```

```
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  490..1695
US-08-643-563A-20


  Query Match             10.8%;  Score 28;  DB 1;  Length 1723;
  Best Local Similarity   47.0%;  Pred. No. 2.2;
  Matches   47;  Conservative    8;  Mismatches   45;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
          || :: |  | ||:||:||| |   ||||:  |  | |   ||| |  |   |  ||| |
Db    808 AGCTTCGTTAACATGGTGGAGCGAGACCGTGCCCTGGGCCACCAGGAGCCCCATTGGAAG 867

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugg 100
            |: ||   : |  |   | || |  ||| |: ||| ||
Db    868 GAGTTCCGCTTTGACCTGACCCAGATCCCGGCTGGGGAGG 907


RESULT  14
US-08-447-570-28
; Sequence 28, Application US/08447570
; Patent No. 5714589
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  OSTEOGENIC DEVICES
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  U.S.A.
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/447,570
;      FILING DATE:  21-FEB-1992
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 810,560
;      FILING DATE:  20-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 827,052
;      FILING DATE:  28-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 660,162
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,988
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,849
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 616,374
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 600,024
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 599,543
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 579,865
;      FILING DATE:  07-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 569,920
;      FILING DATE:  20-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 483,913
;      FILING DATE:  22-FEB-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 422,613
;      FILING DATE:  17-OCT-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 315,342
;      FILING DATE:  23-FEB-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 232,630
;      FILING DATE:  15-AUG-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 179,460
;      FILING DATE:  08-APR-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER, EDMUND R.
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-001CP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1723 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  490..1696
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "hOP2-PP"
;      OTHER INFORMATION:  /note= "hOP2 (cDNA)"
US-08-447-570-28


  Query Match             10.8%;  Score 28;  DB 1;  Length 1723;
  Best Local Similarity   47.0%;  Pred. No. 2.2;
  Matches   47;  Conservative    8;  Mismatches   45;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
          || :: |  | ||:||:||| |   ||||:  |  | |   ||| |  |   |  ||| |
Db    808 AGCTTCGTTAACATGGTGGAGCGAGACCGTGCCCTGGGCCACCAGGAGCCCCATTGGAAG 867

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugg 100
            |: ||   : |  |   | || |  ||| |: ||| ||
Db    868 GAGTTCCGCTTTGACCTGACCCAGATCCCGGCTGGGGAGG 907


RESULT  15
US-08-643-763A-20
; Sequence 20, Application US/08643763A
; Patent No. 5733878
;  GENERAL INFORMATION:
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  OPPERMANN, HERMAN
;    APPLICANT:  COHEN, CHARLES M.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  MORPHOGENIC-INDUCED PERIODONTAL TISSUE
;    TITLE OF INVENTION:  REGENERATION.
```

```
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/643,763A
;      FILING DATE:  06-MAY-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FENTON Esq., GILLIAN M.
;      REGISTRATION NUMBER:  36,508
;      REFERENCE/DOCKET NUMBER:  CRP-067CN
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7560
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1723 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  490..1695
US-08-643-763A-20


  Query Match             10.8%;  Score 28;  DB 1;  Length 1723;
  Best Local Similarity   47.0%;  Pred. No. 2.2;
  Matches   47;  Conservative    8;  Mismatches   45;  Indels    0;  Gaps    0;

Qy       1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccaugucugaaggggaug 60
           || :: |  | ||:||:||| |   ||||:  | || |   ||| |  |   |  ||| |
Db     808 AGCTTCGTTAACATGGTGGAGCGAGACCGTGCCCTGGGCCACCAGGAGCCCCATTGGAAG 867

Qy      61 acgucccuucuggcucauccacaaaaaccgucucgggugg 100
             |: ||   : |  |   | || |  ||| |: ||| ||
Db     868 GAGTTCCGCTTTGACCTGACCCAGATCCCGGCTGGGGAGG 907



Search completed: February 16, 2002, 16:16:56
Job time: 3739 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: February 16, 2002, 15:22:07 ; Search time 2783.16 Seconds
(without alignments)
1631.001 Million cell updates/sec

Title: US-09-587-653-1
Perfect score: 259
Sequence: 1 aguuuggcgaccauggugga.....cccgcuuggaauuaaaaacu 259

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 17159718 seqs, 8763200856 residues

Total number of hits satisfying chosen parameters: 34319436

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US093_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US098_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US099_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		259	100.0	259	22	US-09-587-653-1	Sequence 1, Appli
2		259	100.0	9399	22	US-09-587-653-2	Sequence 2, Appli
3		259	100.0	9399	29	US-09-742-659-1	Sequence 1, Appli
4		259	100.0	9399	46	US-60-137-694-1	Sequence 1, Appli
5		257.4	99.4	259	22	US-09-579-302-2	Sequence 2, Appli
6	c	257.4	99.4	259	22	US-09-579-302-3	Sequence 3, Appli
7		257.4	99.4	259	22	US-09-579-302-4	Sequence 4, Appli
8	c	257.4	99.4	259	22	US-09-579-302-5	Sequence 5, Appli
9		257.4	99.4	309	22	US-09-579-302-1	Sequence 1, Appli
10		82	31.7	82	22	US-09-579-302-14	Sequence 14, Appl
11	c	66	25.5	108	22	US-09-587-653-12	Sequence 12, Appl
12	c	35.4	13.7	468	23	US-09-617-681A-4316	Sequence 4316, Ap
13	c	34.4	13.3	790	29	US-09-726-788-3085	Sequence 3085, Ap
14		32.6	12.6	413	17	US-09-359-067-40731	Sequence 40731, A
15		32.2	12.4	429	17	US-09-306-350A-23940	Sequence 23940, A
16		32.2	12.4	429	32	US-09-909-629-23940	Sequence 23940, A
17		32.2	12.4	474	51	US-60-181-996-546	Sequence 546, App
18		32.2	12.4	527	51	US-60-187-385-242	Sequence 242, App
19	c	32.2	12.4	595	50	US-60-173-469-808	Sequence 808, App
20		32.2	12.4	30533	57	US-60-245-221-27	Sequence 27, Appl
21	c	32	12.4	262	14	US-09-031-483-174	Sequence 174, App
22	c	32	12.4	262	20	US-09-535-897-30627	Sequence 30627, A
23	c	32	12.4	262	37	US-60-044-653-174	Sequence 174, App
24	c	32	12.4	3179	50	US-60-176-083-9	Sequence 9, Appli
25	c	32	12.4	7564	65	US-60-324-185-1366	Sequence 1366, Ap
26	c	31.8	12.3	32768	56	US-60-230-445-752	Sequence 752, App
27		31.2	12.0	580	58	US-60-253-378-35168	Sequence 35168, A
28		31.2	12.0	707	28	US-09-717-350-1933	Sequence 1933, Ap
29	c	31	12.0	553	12	US-08-803-609-2555	Sequence 2555, Ap
30	c	31	12.0	553	32	US-09-912-292-17116	Sequence 17116, A
31		31	12.0	6322	24	US-09-637-889-9712	Sequence 9712, Ap
32		31	12.0	6322	24	US-09-637-890-9762	Sequence 9762, Ap
33		31	12.0	6322	25	US-09-644-867-7855	Sequence 7855, Ap
34		31	12.0	6322	25	US-09-644-869-9708	Sequence 9708, Ap
35		31	12.0	6322	25	US-09-652-121-6694	Sequence 6694, Ap
36		31	12.0	6322	25	US-09-652-124-8629	Sequence 8629, Ap
37		31	12.0	6322	25	US-09-652-127-9967	Sequence 9967, Ap
38		31	12.0	6322	25	US-09-652-816-9036	Sequence 9036, Ap
39		31	12.0	6322	26	US-09-667-617-1655	Sequence 1655, Ap
40		31	12.0	6322	27	US-09-698-014-4911	Sequence 4911, Ap
41		31	12.0	6322	28	US-09-710-281-4418	Sequence 4418, Ap

	42	31	12.0	6322	28	US-09-716-475-8060	Sequence 8060, Ap
	43	31	12.0	6322	29	US-09-726-788-6096	Sequence 6096, Ap
	44	31	12.0	6322	29	US-09-726-807-4020	Sequence 4020, Ap
c	45	31	12.0	205153	54	US-60-216-770-14	Sequence 14, Appl

ALIGNMENTS

```
RESULT   1
US-09-587-653-1
; Sequence 1, Application US/09587653
; GENERAL INFORMATION:
;  APPLICANT:  SANGAR, DAVID V.
;  APPLICANT:  LEMON, STANLEY M.
;  TITLE OF INVENTION:  3' SEQUENCE OF THE GB VIRUS B (GBV-B) GENOME
;  FILE REFERENCE:  UTSG:231US
;  CURRENT APPLICATION NUMBER: US/09/587,653
;  CURRENT FILING DATE:  2000-06-05
;  PRIOR APPLICATION NUMBER:  60/137,665
;  PRIOR FILING DATE:  1999-06-04
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 259
;   TYPE: RNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism:  GB VIRUS B
US-09-587-653-1


  Query Match            100.0%;  Score 259;  DB 22;  Length 259;
  Best Local Similarity  100.0%;  Pred. No. 3.7e-74;
  Matches  259;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccaugucugaaggggaug 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccaugucugaaggggaug 60

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120

Qy    121 gggaagcagucaguauaauucccgucgugugguggugacgccucacgacguauuuguccgc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 gggaagcagucaguauaauucccgucgugugguggugacgccucacgacguauuuguccgc 180

Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240

Qy    241 ccgcuuggaauuaaaaacu 259
          ||||||||||||||||||
Db    241 ccgcuuggaauuaaaaacu 259


RESULT   2
US-09-587-653-2
; Sequence 2, Application US/09587653
; GENERAL INFORMATION:
;  APPLICANT:  SANGAR, DAVID V.
;  APPLICANT:  LEMON, STANLEY M.
;  TITLE OF INVENTION:  3' SEQUENCE OF THE GB VIRUS B (GBV-B) GENOME
;  FILE REFERENCE:  UTSG:231US
;  CURRENT APPLICATION NUMBER: US/09/587,653
;  CURRENT FILING DATE:  2000-06-05
;  PRIOR APPLICATION NUMBER:  60/137,665
;  PRIOR FILING DATE:  1999-06-04
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 9399
;   TYPE: DNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism:  GB VIRUS B
US-09-587-653-2


  Query Match            100.0%;  Score 259;  DB 22;  Length 9399;
  Best Local Similarity  76.1%;  Pred. No. 1e-73;
  Matches  197;  Conservative   62;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccaugucugaaggggaug 60
          ||:::|||||||:||:|||:|||||||:::||||:|||||||:||:|:|||||||:|
Db   9141 agtttggcgaccatggtggatcagaaccgtttcgggtgaagccatggtctgaaggggatg 9200

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          |||:|||::|:|||:||:|||||||||||||:|:||||:|||:||||||:||:|||:|:|:
Db   9201 acgtcccttctggctcatccacaaaaaccgtctcgggtgggtgaggagtcctggctgtgt 9260

Qy    121 gggaagcagucaguauaauucccgucgugugguggugacgccucacgacguauuuguccgc 180
          |||||||||:|||:|:||::||||:||:|:|:||:||||||:|||||||:|:::|:||||
Db   9261 gggaagcagtcagtataattcccgtcgtgtgtggtgacgcctcacgacgtatttgtccgc 9320

Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          :|:|||||||:||:|||||||||:|||||||:::::|::|||||||||||||||||||
Db   9321 tgtgcagagcgtagtaccaagggctgcaccccggtttttgttccaagcggagggcaaccc 9380

Qy    241 ccgcuuggaauuaaaaacu 259
          ||||::||||::||||||:
Db   9381 ccgcttggaattaaaaact 9399


RESULT   3
US-09-742-659-1
; Sequence 1, Application US/09742659
; GENERAL INFORMATION:
;  APPLICANT:  Hong, Zhi
;  APPLICANT:  Butkiewicz, Nancy J.
;  APPLICANT:  Zhong, Weidong
;  APPLICANT:  Ingravallo, Paul
;  APPLICANT:  Wright-Minogue, Jacquelyn
;  APPLICANT:  Lau, Johnson Y.
;  APPLICANT:  Lemon, Stanley M.
;  TITLE OF INVENTION:  Chimeric HCV/GBV-B viruses
;  FILE REFERENCE:  ID01116
;  CURRENT APPLICATION NUMBER: US/09/742,659
;  CURRENT FILING DATE:  2000-12-21
;  PRIOR APPLICATION NUMBER:  US 60/171,469
;  PRIOR FILING DATE:  1999-12-22
;  NUMBER OF SEQ ID NOS:  16
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 9399
;   TYPE: DNA
;   ORGANISM: GB virus-B
US-09-742-659-1


  Query Match            100.0%;  Score 259;  DB 29;  Length 9399;
  Best Local Similarity  76.1%;  Pred. No. 1e-73;
  Matches  197;  Conservative   62;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccaugucugaaggggaug 60
          ||:::|||||||:||:|||:|||||||:::||||:|||||||:||:|:|||||||:|
Db   9141 agtttggcgaccatggtggatcagaaccgtttcgggtgaagccatggtctgaaggggatg 9200

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          |||:|||::|:|||:||:|||||||||||||:|:||||:|||:||||||:||:|||:|:|:
Db   9201 acgtcccttctggctcatccacaaaaaccgtctcgggtgggtgaggagtcctggctgtgt 9260

Qy    121 gggaagcagucaguauaauucccgucgugugguggugacgccucacgacguauuuguccgc 180
          |||||||||:|||:|:||::||||:||:|:|:||:||||||:|||||||:|:::|:||||
Db   9261 gggaagcagtcagtataattcccgtcgtgtgtggtgacgcctcacgacgtatttgtccgc 9320
```

```
Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          :|:||||||||:||:|||||||||:|||||||||:::::|::||||||||||||||||||
Db   9321 tgtgcagagcgtagtaccaagggctgcaccccggtttttgttccaagcggagggcaaccc 9380

Qy    241 ccgcuuggaauuaaaaacu 259
          ||||::||||::||||||:
Db   9381 ccgcttggaattaaaaact 9399


RESULT   4
US-60-137-694-1
; Sequence 1, Application US/60137694
; GENERAL INFORMATION:
;  APPLICANT: Bukh, Jens
;  APPLICANT: Yanagi, Masayuki
;  APPLICANT: Purcell, Robert
;  APPLICANT: Emerson, Suzanne
;  TITLE OF INVENTION: Infectious cDNA Clone of GB virus B and Uses Thereof
;  FILE REFERENCE: 20264308
;  CURRENT APPLICATION NUMBER: US/60/137,694
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 9399
;   TYPE: DNA
;   ORGANISM: GBV-B virus
US-60-137-694-1


  Query Match             100.0%;  Score 259;  DB 46;  Length 9399;
  Best Local Similarity    76.1%;  Pred. No. 1e-73;
  Matches  197;  Conservative   62;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
          ||:::|||||||||:||:|||:|||||||||:::||||:|||||||:||:|:|||||||||:|
Db   9141 agtttggcgaccatggtggatcagaaccgtttcgggtgaagccatggtctgaaggggatg 9200

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          |||:|||::|:|||:||:||||||||||||:|:||||:|||:||||||:||:|||:|:|:
Db   9201 acgtcccttctggctcatccacaaaaaccgtctcgggtgggtgaggagtcctggctgtgt 9260

Qy    121 gggaagcagucaguauaaauucccgucgugugguggugacgccucacgacguauuuguccgc 180
          |||||||||:|||:|:||::||||:||:|:|:||||||||:|||||||:|:::|:||||
Db   9261 gggaagcagtcagtataattcccgtcgtgtgtggtgacgcctcacgacgtatttgtccgc 9320

Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          :|:||||||||:||:|||||||||:|||||||||:::::|::||||||||||||||||||
Db   9321 tgtgcagagcgtagtaccaagggctgcaccccggtttttgttccaagcggagggcaaccc 9380

Qy    241 ccgcuuggaauuaaaaacu 259
          ||||::||||::||||||:
Db   9381 ccgcttggaattaaaaact 9399


RESULT   5
US-09-579-302-2
; Sequence 2, Application US/09579302
; GENERAL INFORMATION:
;  APPLICANT: Traboni, Cinzia
;  TITLE OF INVENTION: Novel GBV sequence
;  FILE REFERENCE: Mewburn
;  CURRENT APPLICATION NUMBER: US/09/579,302
;  CURRENT FILING DATE:  2000-05-25
;  PRIOR APPLICATION NUMBER: GB 9912432.3
;  PRIOR FILING DATE: 1999-05-27
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 259
;   TYPE: RNA
;   ORGANISM: GBV-B-like virus
US-09-579-302-2


  Query Match             99.4%;  Score 257.4;  DB 22;  Length 259;
  Best Local Similarity   99.6%;  Pred. No. 1.2e-73;
  Matches  258;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120

Qy    121 gggaagcagucaguauaaauucccgucgugugguggugacgccucacgacguauuuguccgc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 gggaagcagucaguauaaauucccgucgugugguggugacgccucacgacguacuuguccgc 180

Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240

Qy    241 ccgcuuggaauuaaaaacu 259
          |||||||||||||||||||
Db    241 ccgcuuggaauuaaaaacu 259


RESULT   6
US-09-579-302-3/c
; Sequence 3, Application US/09579302
; GENERAL INFORMATION:
;  APPLICANT: Traboni, Cinzia
;  TITLE OF INVENTION: Novel GBV sequence
;  FILE REFERENCE: Mewburn
;  CURRENT APPLICATION NUMBER: US/09/579,302
;  CURRENT FILING DATE:  2000-05-25
;  PRIOR APPLICATION NUMBER: GB 9912432.3
;  PRIOR FILING DATE: 1999-05-27
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 259
;   TYPE: RNA
;   ORGANISM: GBV-B-like virus
US-09-579-302-3


  Query Match             99.4%;  Score 257.4;  DB 22;  Length 259;
  Best Local Similarity   76.1%;  Pred. No. 1.2e-73;
  Matches  197;  Conservative   61;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
          ||:::|||||||||:||:|||:|||||||||:::||||:|||||||:||:|:|||||||||:|
Db    259 AGTTTGGCGACCATGGTGGATCAGAACCGTTTCGGGTGAAGCCATGGTCTGAAGGGGATG 200

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          |||:|||::|:|||:||:||||||||||||:|:||||:|||:||||||:||:|||:|:|:
Db    199 ACGTCCCTTCTGGCTCATCCACAAAAACCGTCTCGGGTGGGTGAGGAGTCCTGGCTGTGT 140

Qy    121 gggaagcagucaguauaaauucccgucgugugguggugacgccucacgacguauuuguccgc 180
          |||||||||:|||:|:||::||||:||:|:|:||||||||:|||||||:|:::|:||||
Db    139 GGGAAGCAGTCAGTATAATTCCCGTCGTGTGTGGTGACGCCTCACGACGTACTTGTCCGC 80

Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          :|:||||||||:||:|||||||||:|||||||||:::::|::||||||||||||||||||
Db     79 TGTGCAGAGCGTAGTACCAAGGGCTGCACCCCGGTTTTTGTTCCAAGCGGAGGGCAACCC 20

Qy    241 ccgcuuggaauuaaaaacu 259
          ||||::||||::||||||:
Db     19 CCGCTTGGAATTAAAAACT 1
```

```
RESULT   7
US-09-579-302-4
; Sequence 4, Application US/09579302
; GENERAL INFORMATION:
;  APPLICANT: Traboni, Cinzia
;  TITLE OF INVENTION: Novel GBV sequence
;  FILE REFERENCE: Mewburn
;  CURRENT APPLICATION NUMBER: US/09/579,302
;  CURRENT FILING DATE:  2000-05-25
;  PRIOR APPLICATION NUMBER: GB 9912432.3
;  PRIOR FILING DATE: 1999-05-27
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 259
;   TYPE: DNA
;   ORGANISM: GBV-B-like virus
;   FEATURE:
;   NAME/KEY: 3'UTR
;   LOCATION: (1)..(259)
;   OTHER INFORMATION: cDNA of positive strand 3'UTR of GBV-B
US-09-579-302-4


  Query Match             99.4%;  Score 257.4;  DB 22;  Length 259;
  Best Local Similarity   76.1%;  Pred. No. 1.2e-73;
  Matches  197;  Conservative   61;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
          ||:::|||||||||:||:|||:||||||||:::||||:|||||||:||:|:|||||||:|
Db      1 agtttggcgaccatggtggatcagaaccgtttcgggtgaagccatggtctgaaggggatg 60

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          |||:|||::|:|||:||:|||||||||||||:|:||||:|||:||||||:||:|||:|:|:
Db     61 acgtcccttctggctcatccacaaaaaccgtctcgggtgggtgaggagtcctggctgtgt 120

Qy    121 gggaagcagucaguauaauucccgucgugugguggugacgccucacgacguauuuguccgc 180
          |||||||||:|||:|:||::||||:||:|:|:||:||||||:|||||||:| ::|:||||
Db    121 gggaagcagtcagtataattcccgtcgtgtgtggtgacgcctcacgacgtacttgtccgc 180

Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          :|:|||||||||:||:|||||||||:|||||||||:::::|::||||||||||||||||||
Db    181 tgtgcagagcgtagtaccaagggctgcaccccggtttttgttccaagcggagggcaaccc 240

Qy    241 ccgcuuggaauuaaaaacu 259
          ||||::||||::|||||||:
Db    241 ccgcttggaattaaaaact 259


RESULT   8
US-09-579-302-5/c
; Sequence 5, Application US/09579302
; GENERAL INFORMATION:
;  APPLICANT: Traboni, Cinzia
;  TITLE OF INVENTION: Novel GBV sequence
;  FILE REFERENCE: Mewburn
;  CURRENT APPLICATION NUMBER: US/09/579,302
;  CURRENT FILING DATE:  2000-05-25
;  PRIOR APPLICATION NUMBER: GB 9912432.3
;  PRIOR FILING DATE: 1999-05-27
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 259
;   TYPE: DNA
;   ORGANISM: GBV-B-like virus
;   FEATURE:
;   NAME/KEY: 3'UTR
;   LOCATION: (1)..(259)
;   OTHER INFORMATION: cDNA of negative strand 3'UTR of GBV-B
US-09-579-302-5


  Query Match             99.4%;  Score 257.4;  DB 22;  Length 259;
  Best Local Similarity   76.1%;  Pred. No. 1.2e-73;
  Matches  197;  Conservative   61;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
          ||:::|||||||||:||:|||:||||||||:::||||:|||||||:||:|:|||||||:|
Db    259 AGTTTGGCGACCATGGTGGATCAGAACCGTTTCGGGTGAAGCCATGGTCTGAAGGGGATG 200

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          |||:|||::|:|||:||:|||||||||||||:|:||||:|||:||||||:||:|||:|:|:
Db    199 ACGTCCCTTCTGGCTCATCCACAAAAACCGTCTCGGGTGGGTGAGGAGTCCTGGCTGTGT 140

Qy    121 gggaagcagucaguauaauucccgucugugugguggugacgccucacgacguauuuguccgc 180
          |||||||||:|||:|:||::||||:||:|:|:||:||||||:|||||||:| ::|:||||
Db    139 GGGAAGCAGTCAGTATAATTCCCGTCGTGTGTGGTGACGCCTCACGACGTACTTGTCCGC 80

Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          :|:|||||||||:||:|||||||||:|||||||||:::::|::||||||||||||||||||
Db     79 TGTGCAGAGCGTAGTACCAAGGGCTGCACCCCGGTTTTTGTTCCAAGCGGAGGGCAACCC 20

Qy    241 ccgcuuggaauuaaaaacu 259
          ||||::||||::|||||||:
Db     19 CCGCTTGGAATTAAAAACT 1


RESULT   9
US-09-579-302-1
; Sequence 1, Application US/09579302
; GENERAL INFORMATION:
;  APPLICANT: Traboni, Cinzia
;  TITLE OF INVENTION: Novel GBV sequence
;  FILE REFERENCE: Mewburn
;  CURRENT APPLICATION NUMBER: US/09/579,302
;  CURRENT FILING DATE:  2000-05-25
;  PRIOR APPLICATION NUMBER: GB 9912432.3
;  PRIOR FILING DATE: 1999-05-27
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 309
;   TYPE: DNA
;   ORGANISM: GBV-B-like virus
;   FEATURE:
;   NAME/KEY: 3'UTR
;   LOCATION: (1)..(309)
;   OTHER INFORMATION: cDNA complementary to the last 49 published 3'UTR
;   OTHER INFORMATION: nucleotides of GBV-B plus a novel nucleotide (C)
;   OTHER INFORMATION: in position 44 and 259 novel nucleotide sequence
;   OTHER INFORMATION: at the 3' end.
US-09-579-302-1


  Query Match             99.4%;  Score 257.4;  DB 22;  Length 309;
  Best Local Similarity   76.1%;  Pred. No. 1.3e-73;
  Matches  197;  Conservative   61;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
          ||:::|||||||||:||:|||:||||||||:::||||:|||||||:||:|:|||||||:|
Db     51 agtttggcgaccatggtggatcagaaccgtttcgggtgaagccatggtctgaaggggatg 110

Qy     61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
          |||:|||::|:|||:||:|||||||||||||:|:||||:|||:||||||:||:|||:|:|:
Db    111 acgtcccttctggctcatccacaaaaaccgtctcgggtgggtgaggagtcctggctgtgt 170

Qy    121 gggaagcagucaguauaauucccgucugugugguggugacgccucacgacguauuuguccgc 180
          |||||||||:|||:|:||::||||:||:|:|:||:||||||:|||||||:| ::|:||||
Db    171 gggaagcagtcagtataattcccgtcgtgtgtggtgacgcctcacgacgtacttgtccgc 230

Qy    181 ugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaaccc 240
          :|:|||||||||:||:|||||||||:|||||||||:::::|::||||||||||||||||||
Db    231 tgtgcagagcgtagtaccaagggctgcaccccggtttttgttccaagcggagggcaaccc 290

Qy    241 ccgcuuggaauuaaaaacu 259
```

```
          ||||::|||||::||||||:
Db    291 ccgcttggaattaaaaact 309
```

NOT INTERF

```
RESULT  10
US-09-579-302-14
; Sequence 14, Application US/09579302
; GENERAL INFORMATION:
;  APPLICANT: Traboni, Cinzia
;  TITLE OF INVENTION: Novel GBV sequence
;  FILE REFERENCE: Mewburn
;  CURRENT APPLICATION NUMBER: US/09/579,302
;  CURRENT FILING DATE:  2000-05-25
;  PRIOR APPLICATION NUMBER: GB 9912432.3
;  PRIOR FILING DATE: 1999-05-27
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
;   LENGTH: 82
;   TYPE: RNA
;   ORGANISM: GBV-B-like virus
US-09-579-302-14


  Query Match             31.7%;  Score 82;  DB 22;  Length 82;
  Best Local Similarity   100.0%;  Pred. No. 3e-16;
  Matches   82;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    178 cgcugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaa 237
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 cgcugugcagagcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaa 60

Qy    238 cccccgcuuggaauuaaaaacu 259
          |||||||||||||||||||||||
Db     61 cccccgcuuggaauuaaaaacu 82
```

```
RESULT  11
US-09-587-653-12/c
; Sequence 12, Application US/09587653
; GENERAL INFORMATION:
;  APPLICANT:  SANGAR, DAVID V.
;  APPLICANT:  LEMON, STANLEY M.
;  TITLE OF INVENTION:  3' SEQUENCE OF THE GB VIRUS B (GBV-B) GENOME
;  FILE REFERENCE:  UTSG:231US
;  CURRENT APPLICATION NUMBER: US/09/587,653
;  CURRENT FILING DATE:  2000-06-05
;  PRIOR APPLICATION NUMBER:  60/137,665
;  PRIOR FILING DATE:  1999-06-04
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 108
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  Synthetic Primer
US-09-587-653-12


  Query Match             25.5%;  Score 66;  DB 22;  Length 108;
  Best Local Similarity   78.8%;  Pred. No. 5.7e-11;
  Matches   52;  Conservative   14;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    194 guaccaagggcugcaccccgguuuuuguuccaagcggagggcaacccccgcuuggaauua 253
          |:|||||||||:|||||||||:::::|::|||||||||||||||||||||::|||||::|
Db    108 GTACCAAGGGCTGCACCCCGGTTTTTGTTCCAAGCGGAGGGCAACCCCCGCTTGGAATTA 49

Qy    254 aaaacu 259
          ||||||
Db     48 AAAACT 43
```

```
RESULT  12
US-09-617-681A-4316/c
; Sequence 4316, Application US/09617681A
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptid
;  TITLE OF INVENTION:  Thereby
;  FILE REFERENCE: 2750-1064P
;  CURRENT APPLICATION NUMBER: US/09/617,681A
;  CURRENT FILING DATE:  2000-07-19
;  NUMBER OF SEQ ID NOS: 9311
; SEQ ID NO 4316
;   LENGTH: 468
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..468
;   OTHER INFORMATION: Ceres Seq. ID 1433311
;   NAME/KEY: misc_feature
;   LOCATION: 1..468
;   OTHER INFORMATION: any n = a, g, c, t, unknown, or other
US-09-617-681A-4316


  Query Match             13.7%;  Score 35.4;  DB 23;  Length 468;
  Best Local Similarity   39.1%;  Pred. No. 0.94;
  Matches   59;  Conservative   20;  Mismatches   72;  Indels    0;  Gaps    0;

Qy     32 ucgggugaagccauggucugaaggggaugacgucccuucuggcucauccacaaaaaccgu 91
          :| | :|  | ||:||  :      ||:|| | || : | | | ||:|| |     || :
Db    206 TCAGCTGCTGTCATGGCATTGGCATGATGATGCCCGTGCCGCCCCATCCTCCCTCTCCCT 147

Qy     92 cucgggugggugaggaguccuggcuguguggggaagcagucaguauaauucccgucgugug 151
          |   ||   | ||| |  : :|||  : :|||| |  : ||   | ::|    || |
Db    146 CGTCGGACCGCGAGAAAGSGTGGCGCTTTGGGATCCCTTTAGAGCAGTTCAATGCGAGCT 87

Qy    152 uggugacgccucacgacguauuuguccgcug 182
          :||     | :| || |   : :|  | |:|
Db     86 TGGAATAACGTCTCGGCACGTCTGCACTCTG 56
```

```
RESULT  13
US-09-726-788-3085/c
; Sequence 3085, Application US/09726788
; GENERAL INFORMATION:
;  APPLICANT: Gearing, David P.
;  APPLICANT:  Kingsbury, Gillian A.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
;  TITLE OF INVENTION:  THEREFOR
;  FILE REFERENCE: 1600.2051-001
;  CURRENT APPLICATION NUMBER: US/09/726,788
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: 60/168,131
;  PRIOR FILING DATE: 1999-11-30
;  NUMBER OF SEQ ID NOS: 7691
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3085
;   LENGTH: 790
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(790)
;   OTHER INFORMATION: n = A,T,C or G
US-09-726-788-3085


  Query Match             13.3%;  Score 34.4;  DB 29;  Length 790;
  Best Local Similarity   45.1%;  Pred. No. 2.3;
  Matches   64;  Conservative   10;  Mismatches   68;  Indels    0;  Gaps    0;

Qy     46 ggucugaaggggaugacgucccuucuggcucauccacaaaaaccgucucgggugggugag 105
```

```
          || |  || ||||:    :|||::|:  |: |    ||||   | :|    | |||:
Db    462 GGCCCCAAAGGGATCGGTTCCCTTCTCACTTACGGCCAAACTTCNTCATTCGGGGGTNTT 403

Qy    106 gaguccuggcuguguggggaagcagucaguauaauuccccgucgugugguggugacgccucac 165
          || || ||  |  |||||   |:  | | ||  | ||    | | || ||| |
Db    402 CAGGCCCGGAAGACCGGGAAAGCGTTGGAAAAAGGCTCGAGACGGGCGGCGACTACGACA 343

Qy    166 gacguauuuguccgcugugcag 187
          |    |   | |||| |:|| |
Db    342 GCTTGAGAAGCCCGCAGTGCGG 321


RESULT 14
US-09-359-067-40731
; Sequence 40731, Application US/09359067
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-747CON1
;  CURRENT APPLICATION NUMBER: US/09/359,067
;  CURRENT FILING DATE: 1999-07-22
;  EARLIER APPLICATION NUMBER: US 09/131,598
;  EARLIER FILING DATE: 1998-08-10
;  NUMBER OF SEQ ID NOS: 49786
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 40731
;   LENGTH: 413
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-359-067-40731


 Query Match            12.6%;  Score 32.6;  DB 17;  Length 413;
 Best Local Similarity  32.5%;  Pred. No. 7.5;
 Matches   62;  Conservative   30;  Mismatches   99;  Indels    0;  Gaps    0;

Qy      3 uuuggcgaccaugguggaucagaaccguuucgggugaagccaugucugaaggggaugac 62
          :::| | |   :  :  |:   |    ::     :| || |    :|:||| |  |: ||
Db    171 tttgacaatgctcattaatattatttttttctatttgtagaccatatctgaaagaaataac 230

Qy     63 gucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugugg 122
          :   ::  |||:|  ||||| | ||  : :|  |   |   |      :|  :|:|:|
Db    231 atttttttaaggctctaccacatagacaatatcatgctagaatgtgtgtgtgtgtgtgtgt 290

Qy    123 gaagcagucaguauaauuccccgucgugugguggugacgccucacgacguauuuguccgcug 182
          |     |:  |:|:|  :|  |  | | |  : |:|  |      ||| :|:  |   || |
Db    291 gtgtatgtatgtataggtcggggagaggatagtggtgggaacagacatataaggaagcgg 350

Qy    183 ugcagagcgua 193
          |  ||  | |
Db    351 ggaggactgga 361


RESULT 15
US-09-306-350A-23940
; Sequence 23940, Application US/09306350A
; GENERAL INFORMATION:
;  APPLICANT:  Dickson, Mark C.
;  APPLICANT:  Drmanac, Radoje T.
;  APPLICANT:  Jones, Lee W.
;  APPLICANT:  Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  TITLE OF INVENTION:  Novel Nucleic Acid Sequences Obtained
;  TITLE OF INVENTION:  From Various cDNA Libraries
;  FILE REFERENCE:  20411-776
;  CURRENT APPLICATION NUMBER: US/09/306,350A
;  CURRENT FILING DATE:  1999-05-07
;  NUMBER OF SEQ ID NOS:  41304
;  SOFTWARE:  pt_CT_1 Version 1.1
; SEQ ID NO 23940
;   LENGTH: 429
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-306-350A-23940


  Query Match            12.4%;  Score 32.2;  DB 17;  Length 429;
  Best Local Similarity  44.4%;  Pred. No. 10;
  Matches   52;  Conservative   12;  Mismatches   53;  Indels    0;  Gaps    0;

Qy      6 ggcgaccaugguggaucagaaccguuucgggugaagccaugucugaaggggaugacguc 65
          | ||   |:| : |  |||||| |  : |  :|||  |:||   ||||  || ||  :
Db    116 gacgtgaatgcttgggcagaacagactggcatgaaggaatggctggaagaagaagatctt 175

Qy     66 ccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugugg 122
          |     ||:||   |  |||| | : :|  | ||     |  :|  |||:|  :||
Db    176 tcgaagagctcaaaaaggaaaatcctttcttgaggcaagagcttctgggctggctgg 232


Search completed: February 16, 2002, 17:03:31
Job time: 6084 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: February 16, 2002, 15:47:52 ; Search time 438.55 Seconds
(without alignments)
1487.656 Million cell updates/sec

Title: US-09-587-653-1
Perfect score: 259
Sequence: 1 aguuuggcgaccauggugga.....cccgcuuggaauuaaaaacu 259

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2186606 seqs, 1259481600 residues

Total number of hits satisfying chosen parameters: 4373212

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq1:*
7: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq2:*
8: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	32.6	12.6	16943	8	US-10-015-127-485	Sequence 485, App
	2	31	12.0	1095	8	US-10-000-897-36	Sequence 36, Appl
	3	30	11.6	381	6	US-09-849-526A-10014	Sequence 10014, A
	4	30	11.6	390	7	US-09-975-673A-5933	Sequence 5933, Ap
	5	30	11.6	49999	6	US-09-647-377-16	Sequence 16, Appl
c	6	29.8	11.5	503	5	US-09-873-402A-78798	Sequence 78798, A
c	7	29.8	11.5	533	5	US-09-873-402A-9421	Sequence 9421, Ap
c	8	29.8	11.5	702	7	US-09-865-439A-74213	Sequence 74213, A
	9	29.4	11.4	385	7	US-09-669-817A-38740	Sequence 38740, A
	10	29.4	11.4	460	8	US-10-029-386-25170	Sequence 25170, A
c	11	29.4	11.4	642	7	US-09-865-439A-79096	Sequence 79096, A
c	12	29.2	11.3	148	6	US-09-963-299-95	Sequence 95, Appl
c	13	29.2	11.3	31782	6	US-09-815-264-72637	Sequence 72637, A
	14	29	11.2	36991	6	US-09-815-264-71589	Sequence 71589, A
c	15	29	11.2	166649	8	US-10-021-698-704	Sequence 704, App
	16	29	11.2	169752	8	US-10-021-698-703	Sequence 703, App
	17	28.8	11.1	348	7	US-09-669-817A-25063	Sequence 25063, A
	18	28.8	11.1	549	8	US-10-021-698-3838	Sequence 3838, Ap
	19	28.8	11.1	891	7	US-09-708-427-18324	Sequence 18324, A
	20	28.6	11.0	393	7	US-09-925-564-6641	Sequence 6641, Ap
	21	28.6	11.0	452	7	US-09-865-419A-22294	Sequence 22294, A
	22	28.6	11.0	1236	7	US-09-708-427-2952	Sequence 2952, Ap
	23	28.4	11.0	3384	9	US-60-341-261-1331	Sequence 1331, Ap
	24	28.4	11.0	5384	9	US-60-341-261-331	Sequence 331, App
c	25	28.2	10.9	347	6	US-09-927-875A-978	Sequence 978, App
	26	28.2	10.9	1501	6	US-09-815-264-19757	Sequence 19757, A
c	27	28.2	10.9	5493	6	US-09-760-446A-2499	Sequence 2499, Ap
c	28	28.2	10.9	5717	6	US-09-760-446A-2500	Sequence 2500, Ap
	29	28.2	10.9	32372	6	US-09-815-264-58867	Sequence 58867, A
c	30	28	10.8	362	7	US-09-669-817A-22274	Sequence 22274, A
c	31	28	10.8	414	7	US-09-675-784A-2905	Sequence 2905, Ap
	32	28	10.8	1723	5	US-09-148-925A-28	Sequence 28, Appl
	33	28	10.8	1723	6	US-09-597-517-20	Sequence 20, Appl
c	34	28	10.8	5650	6	US-09-922-279A-2074	Sequence 2074, Ap
c	35	28	10.8	16440	6	US-09-815-264-59843	Sequence 59843, A
c	36	28	10.8	36711	6	US-09-815-264-63142	Sequence 63142, A
c	37	28	10.8	169998	1	PCT-US01-30551-24	Sequence 24, Appl
c	38	27.8	10.7	229	6	US-09-985-678-77775	Sequence 77775, A
	39	27.8	10.7	389	6	US-09-933-524A-86892	Sequence 86892, A
	40	27.8	10.7	400	6	US-09-985-678-98738	Sequence 98738, A
	41	27.8	10.7	588	7	US-09-865-439A-81324	Sequence 81324, A
	42	27.8	10.7	609	5	US-09-873-402A-84520	Sequence 84520, A
	43	27.8	10.7	609	8	US-10-034-934-43	Sequence 43, Appl
c	44	27.8	10.7	839	7	US-09-865-439A-20253	Sequence 20253, A
	45	27.8	10.7	1498	6	US-09-898-888A-14368	Sequence 14368, A

ALIGNMENTS

RESULT 1
US-10-015-127-485/c
; Sequence 485, Application US/10015127
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Sphingomonas elodea genome sequences and uses thereof
; FILE REFERENCE: 38-10(15806)B
; CURRENT APPLICATION NUMBER: US/10/015,127
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/252,455
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 14357
; SEQ ID NO 485
; LENGTH: 16943
; TYPE: DNA
; ORGANISM: Sphingomonas elodea
US-10-015-127-485

Query Match 12.6%; Score 32.6; DB 8; Length 16943;
Best Local Similarity 50.6%; Pred. No. 2.8;
Matches 40; Conservative 10; Mismatches 29; Indels 0; Gaps 0;

```
Qy      56 ggaugacgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggc 115
           ||  | ||:||  :|  | :   ||| |  | ||| |:   |: | :||||||  |  ||
Db    9744 GGCAGGCGTCCGGTCCAGGTGCGCCAGATCAGCCGGCTGATGTCGATGAGGAGGTCGCGC 9685

Qy     116 ugugugggaagcagucagu 134
           :|  :|||| |||:|||:
Db    9684 TGCATGGGAGCCAGTCAGT 9666
```

RESULT 2
US-10-000-897-36
; Sequence 36, Application US/10000897
; GENERAL INFORMATION:
; APPLICANT: Schueler, P.
; TITLE OF INVENTION: METHODS AND REAGENTS FOR IDENTIFYING RARE FETAL CELLS IN THE
; TITLE OF INVENTION: MATERNAL CIRCULATION
; FILE REFERENCE: 11012-004-999
; CURRENT APPLICATION NUMBER: US/10/000,897
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 60/248,882
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 78

```
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 36
;   LENGTH: 1095
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-000-897-36


  Query Match            12.0%;  Score 31;  DB 8;  Length 1095;
  Best Local Similarity   32.5%;  Pred. No. 4.3;
  Matches   62;  Conservative   29;  Mismatches  100;  Indels    0;  Gaps    0;

Qy      3 uuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaugac 62
          :::| | |   :  :  |:   |    ::     :| || |    :|:||| |  |: ||
Db    620 tttgacaatgctcattaatattatttttctatttgtagaccatatctgaaagaaataac 679

Qy     63 gucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugugg 122
           :   ::  |||:|  ||||| | ||  : :|  |   |   |      :|  :|:|:|
Db    680 attttttaaggctctaccacatagacaatatcatgctagaatgtgtgtgtgtgtgtgtgt 739

Qy    123 gaagcagucaguauaauucccgucgugugguggugacgccucacgacguauuuguccgcug 182
          |     |:  |:|:|  :|  |  | |  : |:|  |      |||  |:  |   || |
Db    740 gtgtatgtatgtataggtcggggagaggatagtggtgggaacagacaaataaggaagcgg 799

Qy    183 ugcagagcgua 193
          |  ||  | |
Db    800 ggaggactgga 810


RESULT   3
US-09-849-526A-10014
; Sequence 10014, Application US/09849526A
; GENERAL INFORMATION:
;  APPLICANT: Ruff, Thomas G.
;  APPLICANT:  Shukla, Hridayabhiranjan
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 16517.250[38-21(51930)B]
;  CURRENT APPLICATION NUMBER: US/09/849,526A
;  CURRENT FILING DATE:  2001-05-07
;  PRIOR APPLICATION NUMBER: US 60/202,214
;  PRIOR FILING DATE: 2000-05-08
;  PRIOR APPLICATION NUMBER: US 09/654,617
;  PRIOR FILING DATE: 2000-09-05
;  PRIOR APPLICATION NUMBER: US 09/684,016
;  PRIOR FILING DATE: 2000-10-10
;  PRIOR APPLICATION NUMBER: US 09/816,660
;  PRIOR FILING DATE: 2001-03-26
;  NUMBER OF SEQ ID NOS: 30131
; SEQ ID NO 10014
;   LENGTH: 381
;   TYPE: DNA
;   ORGANISM: Zea mays subsp. mexicana
;   FEATURE:
;   OTHER INFORMATION: Clone ID: uC-zmroteosinte065h08a1
US-09-849-526A-10014


  Query Match            11.6%;  Score 30;  DB 6;  Length 381;
  Best Local Similarity   38.1%;  Pred. No. 6.9;
  Matches   45;  Conservative   18;  Mismatches   55;  Indels    0;  Gaps    0;

Qy    129 gucaguauaauucccgucgugugguggugacgccucacgacguauuuguccgcugugcaga 188
          | || :|:||  ||||  | |:  ||   | |   |  | :|::: : | | |  ||
Db    175 gccaatataaggcccggggcgtcgggctgcccaaaaaaaaatatttttgcacagcccaac 234

Qy    189 gcguaguaccaagggcugcaccccgguuuuuguuccaagcggagggcaacccccgcuu 246
            :  :    || ||: | ||||| : ::  ::   | ||  ||||||  ||   ::
Db    235 cattgttggggagagctacccccgttgttaatttttaacgcggggcaagtccgtttt 292


RESULT   4
US-09-975-673A-5933
; Sequence 5933, Application US/09975673A
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
;  TITLE OF INVENTION:  From Various Libraries
;  FILE REFERENCE: 786
;  CURRENT APPLICATION NUMBER: US/09/975,673A
;  CURRENT FILING DATE:  2001-10-11
;  PRIOR APPLICATION NUMBER: PRIOR APPLICATIO NUMBER:09/496,911
;  PRIOR FILING DATE: PRIORFILING DATE: 2000-02-03
;  NUMBER OF SEQ ID NOS: 13944
;  SOFTWARE: Hy-patent.pl Version 3.1
; SEQ ID NO 5933
;   LENGTH: 390
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(390)
;   OTHER INFORMATION: n = A,T,C or G
US-09-975-673A-5933


  Query Match            11.6%;  Score 30;  DB 7;  Length 390;
  Best Local Similarity   45.5%;  Pred. No. 6.9;
  Matches   51;  Conservative    9;  Mismatches   52;  Indels    0;  Gaps    0;

Qy    106 gaguccuggcuguguggggaagcagucaguauaauucccgucgugugguggugacgccucac 165
          | |:|| |||:| | |   ||||  || :|  | :   | | : :  || | |    | |
Db    212 gcgtccaggctgcgcgatgagcacacactactactgatggcctttcgggggccctgcccc 271

Qy    166 gacguauuuguccgcugugcagagcguaguaccaagggcugcaccccgguuu 217
            ||      | ||| :|  ||| ||   | ||    ||  ||| |||||  ::
Db    272 aaccggagaggccggtgcacagggccccgcacgtnnggggcatcccggctt 323


RESULT   5
US-09-647-377-16
; Sequence 16, Application US/09647377
; GENERAL INFORMATION:
;  APPLICANT: Rosenthal, Andre et al.
;  TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING PROTEINS WHICH INFLUENCE BONE
;  FILE REFERENCE: 0147-0211P
;  CURRENT APPLICATION NUMBER: US/09/647,377
;  CURRENT FILING DATE:  2000-09-24
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
;   LENGTH: 49999
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-647-377-16


  Query Match            11.6%;  Score 30;  DB 6;  Length 49999;
  Best Local Similarity   39.8%;  Pred. No. 27;
  Matches   47;  Conservative   16;  Mismatches   55;  Indels    0;  Gaps    0;

Qy     14 ugguggaucagaaccguuucgggugaagccauggucugaaggggaugacgucccuucugg 73
          :||:|||:|||| |  :: | | :|  | ||: |:| | ||  || ||  :   :  :|
Db  24459 tggtggatcagagcacttgcagttggggacattgtcagtagtagaagatctgagtgatgc 24518

Qy     74 cucauccacaaaaaccgucucgggugggugaggaguccuggcugugugggaagcaguc 131
          |:| :||           |: | |  || ||     |  ||   :   ||| | ||:|
Db  24519 ctcttcctggctgctgcacttgaggaggggacagagcagggtgttcctggatgaagtc 24576


RESULT   6
US-09-873-402A-78798/c
; Sequence 78798, Application US/09873402A
; GENERAL INFORMATION:
;  APPLICANT: Edgerton, Michael D
;  APPLICANT:  Hardeman, Kristine J.
```

```
;  APPLICANT:  Varagona, Marguerite J.
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(51934)B
;  CURRENT APPLICATION NUMBER: US/09/873,402A
;  CURRENT FILING DATE:  2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/209,830
;  PRIOR FILING DATE: 2000-06-06
;  NUMBER OF SEQ ID NOS: 90966
; SEQ ID NO 78798
;   LENGTH: 503
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3632-059-Q6-N6-C1
US-09-873-402A-78798


  Query Match           11.5%;  Score 29.8;  DB 5;  Length 503;
  Best Local Similarity   47.7%;  Pred. No. 8.7;
  Matches   31;  Conservative   12;  Mismatches   22;  Indels    0;  Gaps    0;

Qy      30 uuucgggugaagccauggucugaaggggaugacgucccuucuggcucauccacaaaaacc 89
           :::|| |:| | |||:|  |:|   |   :  ||:||| :|:|   || | ||||| |
Db      67 TTTCGTGTGTACCCATGAGCTGGTCGCATTCGCGTCCCCTCTGATCCAACAACAAACATT 8

Qy      90 gucuc 94
           |:|:|
Db       7 GTCTC 3


RESULT   7
US-09-873-402A-9421/c
; Sequence 9421, Application US/09873402A
; GENERAL INFORMATION:
;  APPLICANT: Edgerton, Michael D
;  APPLICANT:  Hardeman, Kristine J.
;  APPLICANT:  Varagona, Marguerite J.
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(51934)B
;  CURRENT APPLICATION NUMBER: US/09/873,402A
;  CURRENT FILING DATE:  2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/209,830
;  PRIOR FILING DATE: 2000-06-06
;  NUMBER OF SEQ ID NOS: 90966
; SEQ ID NO 9421
;   LENGTH: 533
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3591-104-Q1-N6-A9
US-09-873-402A-9421


  Query Match           11.5%;  Score 29.8;  DB 5;  Length 533;
  Best Local Similarity   47.7%;  Pred. No. 8.8;
  Matches   31;  Conservative   12;  Mismatches   22;  Indels    0;  Gaps    0;

Qy      30 uuucgggugaagccauggucugaaggggaugacgucccuucuggcucauccacaaaaacc 89
           :::|| |:| | |||:|  |:|   |   :  ||:||| :|:|   || | ||||| |
Db     101 TTTCGTGTGTAACCATGAGCTGGTCGCATTCGCGTCCCCTCTGATCCAACAACAAACATT 42

Qy      90 gucuc 94
           |:|:|
Db      41 GTCTC 37


RESULT   8
US-09-865-439A-74213/c
; Sequence 74213, Application US/09865439A
; GENERAL INFORMATION:
;  APPLICANT: Edgerton, Michael D
;  APPLICANT:  Hardeman, Kristine J.
;  APPLICANT:  La Rosa, Thomas J.
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(51936)B
;  CURRENT APPLICATION NUMBER: US/09/865,439A
;  CURRENT FILING DATE:  2001-05-29
;  PRIOR APPLICATION NUMBER: US 60/207,458
;  PRIOR FILING DATE: 2000-05-30
;  NUMBER OF SEQ ID NOS: 119126
; SEQ ID NO 74213
;   LENGTH: 702
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: jC-zmshLIB3587P031f02a1
US-09-865-439A-74213


  Query Match           11.5%;  Score 29.8;  DB 7;  Length 702;
  Best Local Similarity   47.7%;  Pred. No. 9.5;
  Matches   31;  Conservative   12;  Mismatches   22;  Indels    0;  Gaps    0;

Qy      30 uuucgggugaagccauggucugaaggggaugacgucccuucuggcucauccacaaaaacc 89
           :::|| |:| | |||:|  |:|   |   :  ||:||| :|:|   || | ||||| |
Db     141 TTTCGTGTGTAACCATGAGCTGGTCGCATTCGCGTCCCCTCTGATCCAACAACAAACATT 82

Qy      90 gucuc 94
           |:|:|
Db      81 GTCTC 77


RESULT   9
US-09-669-817A-38740
; Sequence 38740, Application US/09669817A
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  Ruan, Yijun G.
;  APPLICANT:  Wallick, C. Kevin
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(51469)B
;  CURRENT APPLICATION NUMBER: US/09/669,817A
;  CURRENT FILING DATE:  2001-09-26
;  PRIOR APPLICATION NUMBER: US 60/156,951
;  PRIOR FILING DATE: 1999-09-30
;  NUMBER OF SEQ ID NOS: 43701
; SEQ ID NO 38740
;   LENGTH: 385
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: uC-osroM202005h07b1
US-09-669-817A-38740


  Query Match           11.4%;  Score 29.4;  DB 7;  Length 385;
  Best Local Similarity   48.5%;  Pred. No. 11;
  Matches   50;  Conservative    7;  Mismatches   46;  Indels    0;  Gaps    0;

Qy       2 guuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggauga 61
           | ::|| | |||:|| ||| | ||| ||  :   |   ||||  |  |:|| | |||:
Db     133 ggttggtggccatggcggagctgaagcggctgtcggagagccgcgacctgacgcggatcg 192

Qy      62 cgucccuucuggcucauccacaaaaaccgucucgggugggguga 104
             |  | :   ||| ||  | || |  | |   | || |   ||
Db     193 agcgcatcggggcgcactcccacatccgggggctggggctgga 235


RESULT  10
US-10-029-386-25170
; Sequence 25170, Application US/10029386
; GENERAL INFORMATION:
```

```
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25170
;   LENGTH: 460
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL031781.1
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
;   OTHER INFORMATION: SWISSPROT HIT: O00712, EVALUE 6.10e-01
;   OTHER INFORMATION: NT HIT: AB067807.1, EVALUE 0.00e+00
;   OTHER INFORMATION: EST_HUMAN HIT: AV657509.1, EVALUE 0.00e+00
US-10-029-386-25170


  Query Match             11.4%;  Score 29.4;  DB 8;  Length 460;
  Best Local Similarity   38.9%;  Pred. No. 11;
  Matches   79;  Conservative   23;  Mismatches   96;  Indels    5;  Gaps    1;

Qy      35 ggugaagccaugguguagaggggaugacgucccuucuggcucauccacaaaaaccgucuc 94
           || | |  ||: |  :| |  |  : |     |::| ||| |      ||  |  |   |
Db     202 ggggtactcatagggtgtatagattaaaccagcttcgggccctggaggaactattgcagc 261

Qy      95 gggugggugaggaguccuggcuguguggggaagcagucaguauaauuucccguc-----gug 149
             | :|||:|||||||: | |  :|    |  ||| |:| |:|:   :|   :|
Db     262 agttgggtgaggagttccgtttggcatgacagcggtctgtatttgtctgatcaaaggcat 321

Qy     150 uguggugacgccucacgacguauuuguccgcugugcagagcguaguaccaagggcugcac 209
           : :||:   |||    | ||:|   |: ||| | ||||   |: | |  |  |||
Db     322 tatggtagggccagctggcgtaggagtacgcagggcagctggtgggagaaccggcgcagg 381

Qy     210 cccgguuuuuguuccaagcggag 232
           ||| |:  :  : |   | | ||
Db     382 cccagtaatgatccttggagcag 404


RESULT  11
US-09-865-439A-79096/c
; Sequence 79096, Application US/09865439A
; GENERAL INFORMATION:
;  APPLICANT: Edgerton, Michael D
;  APPLICANT:  Hardeman, Kristine J.
;  APPLICANT:  La Rosa, Thomas J.
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(51936)B
;  CURRENT APPLICATION NUMBER: US/09/865,439A
;  CURRENT FILING DATE:  2001-05-29
;  PRIOR APPLICATION NUMBER: US 60/207,458
;  PRIOR FILING DATE: 2000-05-30
;  NUMBER OF SEQ ID NOS: 119126
; SEQ ID NO 79096
;   LENGTH: 642
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3606-049-Q1-K6-D10
US-09-865-439A-79096


  Query Match             11.4%;  Score 29.4;  DB 7;  Length 642;
  Best Local Similarity   37.8%;  Pred. No. 13;
  Matches   42;  Conservative   18;  Mismatches   51;  Indels    0;  Gaps    0;

Qy      82 caaaaaccgucucgggugggugaggaguccuggcuguguggggaagcagucaguauaauuc 141
           |||| |  |:| | |  || : | ||  | :|||  : :  | ||||  |||: :  ::|
Db     259 CAAACAAGGTCCCAGTAGGATCAAGAACCTTGGCCATCTCAGTAGCATACAGTTTGTTTC 200

Qy     142 ccgucgugugguggugacgccucacgacguauuuguccgcugugcagagcgu 192
           || : |:|: :     |  |:||    |  :     |  : :| |||  |:
Db     199 CCATAGTGTATAAATGCAACTCAAACAGGTTACTAGCCTTCTGAAGAAAGT 149


RESULT  12
US-09-963-299-95/c
; Sequence 95, Application US/09963299
; GENERAL INFORMATION:
;  APPLICANT: Abuin, Alejandro
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: Novel Murine Polynucleotide Sequences
;  TITLE OF INVENTION:  and Mutant Cells and Mutant Animals Defined Thereby
;  FILE REFERENCE: LEX-0246-USA
;  CURRENT APPLICATION NUMBER: US/09/963,299
;  CURRENT FILING DATE:  2001-09-26
;  PRIOR APPLICATION NUMBER: US 60/237,272
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 1000
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 95
;   LENGTH: 148
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-963-299-95


  Query Match             11.3%;  Score 29.2;  DB 6;  Length 148;
  Best Local Similarity   41.8%;  Pred. No. 9.7;
  Matches   41;  Conservative   14;  Mismatches   43;  Indels    0;  Gaps    0;

Qy      61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcugugu 120
           || :   ::|: | :||:|   ||||     |:   | |  : || |   | |   | :
Db     102 ACTTTTATTCTTGTTCATCTGTAAAATGGAACTGAAGAGAATAAGAATCTCGGCGAGCTT 43

Qy     121 gggaagcagucaguauaauuucccgucgugugguggugac 158
           ||| || ||:| |:| |     ||:||:|   ||:|||
Db      42 GGGCAGAAGTCTGTAGAGGCGGCGTCGTGAAGGGTGAC 5


RESULT  13
US-09-815-264-72637/c
; Sequence 72637, Application US/09815264
; GENERAL INFORMATION:
;  APPLICANT: Boukharov, Andrey A.
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Dotson, Stanton B.
;  APPLICANT:  Koshi, Jeffrey M.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  APPLICANT:  McIninch, James
;  APPLICANT:  Wu, Wei
;  TITLE OF INVENTION: Genomic Plant Sequences And Uses Thereof
;  FILE REFERENCE: 38-21(51237)G
;  CURRENT APPLICATION NUMBER: US/09/815,264
;  CURRENT FILING DATE:  2001-03-23
;  PRIOR APPLICATION NUMBER: US 09/620,392
;  PRIOR FILING DATE: 2000-07-19
;  PRIOR APPLICATION NUMBER: US 09/702,134
;  PRIOR FILING DATE: 2000-10-31
;  NUMBER OF SEQ ID NOS: 109669
; SEQ ID NO 72637
;   LENGTH: 31782
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-815-264-72637


  Query Match             11.3%;  Score 29.2;  DB 6;  Length 31782;
```

```
  Best Local Similarity   42.8%;  Pred. No. 44;
  Matches   59;  Conservative   11;  Mismatches   68;  Indels    0;  Gaps    0;

Qy      97 gugggugaggaguccuggcuguguugggaagcagucaguauaauucccgucgugugguggug 156
           | |||:  |  |    || :| | | | || |  | |:|    : ||| || |:| ||:
Db   22359 GCGGGTTCGCGGGTTCGGGTGGGGGTGGAGGATCCGGTAAGGCTGCCGCCGCGTGCGGTT 22300

Qy     157 acgccucacgacguauuuguccgcugugcagagcguaguaccaagggcugcaccccgguu 216
             ||     | || | : |:| ||   ||  |||| || |    | ||: |    || ::
Db   22299 GCGGTGGCGGCCGGAGTCGTCGGCGAAGCGAAGCGGAGCAGGCGGAGCTTCGATTCGATT 22240

Qy     217 uuuguuccaagcggaggg 234
             |     |  ||| ||
Db   22239 CGAGGAAGGAAAGGAAGG 22222


RESULT  14
US-09-815-264-71589
; Sequence 71589, Application US/09815264
; GENERAL INFORMATION:
;  APPLICANT: Boukharov, Andrey A.
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Dotson, Stanton B.
;  APPLICANT:  Koshi, Jeffrey M.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  APPLICANT:  McIninch, James
;  APPLICANT:  Wu, Wei
;  TITLE OF INVENTION: Genomic Plant Sequences And Uses Thereof
;  FILE REFERENCE: 38-21(51237)G
;  CURRENT APPLICATION NUMBER: US/09/815,264
;  CURRENT FILING DATE:  2001-03-23
;  PRIOR APPLICATION NUMBER: US 09/620,392
;  PRIOR FILING DATE: 2000-07-19
;  PRIOR APPLICATION NUMBER: US 09/702,134
;  PRIOR FILING DATE: 2000-10-31
;  NUMBER OF SEQ ID NOS: 109669
; SEQ ID NO 71589
;   LENGTH: 36991
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-815-264-71589


  Query Match             11.2%;  Score 29;  DB 6;  Length 36991;
  Best Local Similarity   47.7%;  Pred. No. 53;
  Matches   52;  Conservative    7;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      51 gaaggggaugacgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguc 110
           || | |||:|||| | |  | | |:|     || |   || | |||  | | | ||   |
Db    7861 gaggaggatgacgacgcggcggcctcgcaatcagacgacgacgcggcggagggggggcgc 7920

Qy     111 cuggcugugugggaagcagucaguauaauucccgucgugugguggugacg 159
           | |     |   |  ||||   |  : |: || |: |:|  :|| ||||
Db    7921 cggcggagggacgcggcagagggagtgatgccggttgtggctggggacg 7969


RESULT  15
US-10-021-698-704/c
; Sequence 704, Application US/10021698
; GENERAL INFORMATION:
;  APPLICANT: Tim Keith
;  TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES
;  TITLE OF INVENTION:  RELATING TO RESPIRATORY DISEASES AND OBESITY
;  FILE REFERENCE: HUM01-04
;  CURRENT APPLICATION NUMBER: US/10/021,698
;  CURRENT FILING DATE:  2001-10-22
;  PRIOR APPLICATION NUMBER: 60/211,749
;  PRIOR FILING DATE: 2000-06-14
;  NUMBER OF SEQ ID NOS: 4687
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 704
;   LENGTH: 166649
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(166649)
;   OTHER INFORMATION: n = A,T,C or G
US-10-021-698-704


  Query Match             11.2%;  Score 29;  DB 8;  Length 166649;
  Best Local Similarity   43.6%;  Pred. No. 81;
  Matches   44;  Conservative   12;  Mismatches   45;  Indels    0;  Gaps    0;

Qy       9 gaccaugguggaucagaaccguuucgggugaagccaugucugaaggggaugacgucccu 68
           | || :|  ||        | |: :|     ||||||:||:|    | || :  |   ||
Db   98478 GGCCCTGTGGGTGTTTTCCGGTGTCTTTAAAAGCCATGGTCCAGCGCGGCTTCCTGACCG 98419

Qy      69 ucuggcucauccacaaaaaccgucucgggugggugaggagu 109
             |: | :||:   | ||| ||  |   ||:| |: |||||:
Db   98418 ACTTGTTCATTGGCGAAAGCCAACAGCGGTGAGTCAGGAGT 98378


Search completed: February 16, 2002, 17:12:14
Job time: 5062 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: February 16, 2002, 14:00:57 ; Search time 1527.54 Seconds
(without alignments)
1821.986 Million cell updates/sec

Title: US-09-587-653-1
Perfect score: 259
Sequence: 1 aguuuggcgaccauggugga.....cccgcuuggaauuaaaaacu 259

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	34.8	13.4	670	13	AZ962384	AZ962384 2M0231L11
c	2	33.8	13.1	1092	13	CNS05P0X	AL347370 Tetraodon
	3	33.2	12.8	1508	10	BE615183	BE615183 601281124
c	4	33	12.7	1054	11	BG687472	BG687472 602639444
c	5	32.6	12.6	287	10	AV072177	AV072177 AV072177
c	6	32.6	12.6	561	13	AQ657755	AQ657755 Sheared D
c	7	32	12.4	283	10	BB385246	BB385246 BB385246
	8	31.6	12.2	1022	10	AL568674	AL568674 AL568674
c	9	31.4	12.1	321	10	BB500470	BB500470 BB500470
c	10	31.2	12.0	251	10	BB425694	BB425694 BB425694
c	11	31.2	12.0	281	10	BB421037	BB421037 BB421037
c	12	31.2	12.0	712	11	BF577610	BF577610 602092476
	13	31	12.0	275	10	BE647687	BE647687 UI-M-BH1-
	14	31	12.0	548	10	AW361210	AW361210 RC1-CT025
	15	31	12.0	586	13	BH110789	BH110789 RPCI-24-3
c	16	31	12.0	918	13	CNS02JF4	AL200137 Tetraodon
c	17	31	12.0	1907	12	AK005169	AK005169 Mus muscu
	18	30.6	11.8	469	11	BG639798	BG639798 SD06521.3
	19	30.6	11.8	533	10	AI533127	AI533127 SD04843.5
	20	30.6	11.8	555	10	AI533553	AI533553 SD05422.5
	21	30.6	11.8	654	10	AA951616	AA951616 LD32243.5
	22	30.6	11.8	730	10	AA816991	AA816991 LD21943.5
	23	30.6	11.8	743	10	BE407858	BE407858 601300622
c	24	30.6	11.8	749	13	CNS04ICM	AL292063 Tetraodon
	25	30.6	11.8	1410	11	BG167510	BG167510 602342714
	26	30.4	11.7	259	13	CNS04GUI	AL290115 Tetraodon
c	27	30.4	11.7	278	10	BB365538	BB365538 BB365538
c	28	30.4	11.7	421	10	AU181868	AU181868 AU181868
c	29	30.4	11.7	468	11	BG730062	BG730062 de09c10.y
c	30	30.4	11.7	716	11	BF615851	BF615851 dd97b07.y
c	31	30.4	11.7	750	11	BF613405	BF613405 de01h03.y
	32	30.4	11.7	861	11	BI217114	BI217114 602933448
c	33	30.4	11.7	1135	11	BG336385	BG336385 602405326
c	34	30.2	11.7	279	10	BB382397	BB382397 BB382397
c	35	30.2	11.7	499	11	BG164682	BG164682 602341785
	36	30.2	11.7	1072	13	CNS05C84	AL330781 Tetraodon
c	37	30	11.6	262	10	BB416099	BB416099 BB416099
c	38	30	11.6	314	10	BB162728	BB162728 BB162728
c	39	30	11.6	529	11	BG515211	BG515211 dad66b01.
c	40	30	11.6	546	11	BG512901	BG512901 dad64b05.
c	41	30	11.6	560	11	BG633101	BG633101 GH20211.3
c	42	30	11.6	566	11	BG815751	BG815751 dad72g11.
c	43	30	11.6	575	11	BG512819	BG512819 dad62g08.
	44	30	11.6	655	10	AV703583	AV703583 AV703583
	45	30	11.6	683	10	AV705255	AV705255 AV705255

ALIGNMENTS

RESULT 1
AZ962384
LOCUS AZ962384 670 bp DNA GSS 27-APR-2001
DEFINITION 2M0231L11F Mouse 10kb plasmid UUGC2M library Mus musculus genomic clone UUGC2M0231L11 F, DNA sequence.
ACCESSION AZ962384
VERSION AZ962384.1 GI:13833611
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 670)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0231 row: L column: 11
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 670.
FEATURES Location/Qualifiers
source 1. .670

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0231L11"
/clone_lib="Mouse 10kb plasmid UUGC2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 168 a 151 c 180 g 171 t
ORIGIN

Query Match 13.4%; Score 34.8; DB 13; Length 670;
Best Local Similarity 39.7%; Pred. No. 2.7;
Matches 50; Conservative 19; Mismatches 57; Indels 0; Gaps 0;

```
Qy    133 guauaauucccgucgugugguggugacgccucacgacguauuuguccgcugugcagagcgu 192
          |: : | :|    | :|:|:| :| | ||  ||    :|: : :| ||:   ||    |:
Db    351 GTGTGACTCTGCACATGTGTGTTGTCACCCAACCCTCTATGTCTCTGCTTGCCAACATGT 410

Qy    193 aguaccaagggcugcaccccgguuuuuguuccaagcggagggcaacccccgcuuggaauu 252
          |  |  ||| | :   || | | : ::|:  | |  | | || |||| |   : ||||
Db    411 AAAATGAAGAGATAGCCCTCAGGTGTTGTAACGAATGAATGGAAACCACATGTGGGAAAA 470

Qy    253 aaaaac 258
            | |||
Db    471 GATAAC 476
```

RESULT 2
CNS05P0X/c
LOCUS CNS05P0X 1092 bp DNA GSS 26-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 045G18 of library A from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL347370
VERSION AL347370.1 GI:8241140
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 1092)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Charaterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1092)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1092)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source 1. .1092
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="045G18"
/clone_lib="A"
/note="Genoscope sequence ID : C0AA045BD09C1~end : T7"
BASE COUNT 281 a 240 c 234 g 313 t 24 others
ORIGIN

Query Match 13.1%; Score 33.8; DB 13; Length 1092;
Best Local Similarity 38.6%; Pred. No. 6.1;
Matches 64; Conservative 19; Mismatches 83; Indels 0; Gaps 0;

```
Qy     34 gggugaagccauggucugaaggggaugacgucccuucuggcucauccacaaaaaccgucu 93
          | |:| |||| :  : :||| ||   |  |  |   | | |  | |  | ||  |  :|
Db    654 GAGTGCAGCCTTTCTTTGAATGGAGAGGTGAACAACCAGCCAAACCTGCGAAGGCGTTCA 595

Qy     94 cgggugggugaggaguccuggcuguguggggaagcagucaguauaauuucccgucgugugug 153
          | :||| ||||||    | |: :|:||| ||||| ||| : |    | :|  |   |
Db    594 TCGTTGGGCGAGGAGGGAGGCCTCTGTGGGTAGCAGACAGATTCAGAGGCCTCTGGGCAG 535

Qy    154 gugacgccucacgacguauuuguccgcugugcagagcguaguacca 199
          |    | |    |   : ::: :    :  ||  |   : |:  ||
Db    534 GNTCAGACAACTGGATTTTTTATTTCTTTAGCTAATTATTGTTACA 489
```

RESULT 3
BE615183
LOCUS BE615183 1508 bp mRNA EST 24-AUG-2000
DEFINITION 601281124F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3623261 5', mRNA sequence.
ACCESSION BE615183
VERSION BE615183.1 GI:9896782
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1508)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM299 row: o column: 06
High quality sequence stop: 307.
FEATURES Location/Qualifiers
source 1. .1508
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3623261"
/clone_lib="NIH_MGC_39"
/tissue_type="adenocarcinoma"

/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 316 a 376 c 426 g 388 t 2 others
ORIGIN

Query Match 12.8%; Score 33.2; DB 10; Length 1508;
Best Local Similarity 42.7%; Pred. No. 10;
Matches 47; Conservative 15; Mismatches 48; Indels 0; Gaps 0;

Qy 88 ccgucucgggugggugaggaguccuggcuguguggaagcagucaguauaauuucccgucg 147
Db 1084 CCGCCCCAAATTTTTTTGTGGTGCCGCCCGGGGGGGGGGGGGGGGTATACTCCCCCGTG 1143

Qy 148 uguguggugacgccucacgacguauuuguccgcugugcagagcguaguac 197
Db 1144 GGGGGGGTTTCTCCTCCCGTTGTTGGTGTCCCCTGTGGGGGGGGTACTAC 1193

RESULT 4
BG687472/c
LOCUS BG687472 1054 bp mRNA EST 01-MAY-2001
DEFINITION 602639444F1 NIH_MGC_59 Homo sapiens cDNA clone IMAGE:4762427 5', mRNA sequence.
ACCESSION BG687472
VERSION BG687472.1 GI:13918869
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1054)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLCM1616 row: h column: 12
High quality sequence stop: 336.
FEATURES Location/Qualifiers
source 1. .1054
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4762427"
/clone_lib="NIH_MGC_59"
/tissue_type="mucoepidermoid carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."
BASE COUNT 279 a 316 c 157 g 302 t
ORIGIN

Query Match 12.7%; Score 33; DB 11; Length 1054;
Best Local Similarity 37.3%; Pred. No. 11;
Matches 60; Conservative 21; Mismatches 80; Indels 0; Gaps 0;

Qy 90 gucucgggugggugaggaguccuggcuguguggaagcagucaguauaauuucccgucgug 149
Db 810 GTATGGGAAGTTTGAAGAAAGGTGTGTGTGCGCCATGGAGTTAGGAGGAGTAGTGACCTG 751

Qy 150 uguggugacgccucacgacguauuuguccgcugugcagagcguaguaccaagggcugcac 209
Db 750 TTTGGTGTTGGGCCGCGCAGAAGCCCTGTTGAGCGTAATGTGTAGTAAAAAGGGATGTGG 691

Qy 210 cccgguuuuuguuccaagcggagggcaacccccgcuuggaa 250
Db 690 GAGGGTGTTGTGTGTAAGGAGGAGGCCCGCACACATGGGTA 650

RESULT 5
AV072177/c
LOCUS AV072177 287 bp mRNA EST 24-JUN-1999
DEFINITION AV072177 Mus musculus stomach C57BL/6J adult Mus musculus cDNA clone 2200002K22, mRNA sequence.
ACCESSION AV072177
VERSION AV072177.1 GI:5192005
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 287)
AUTHORS Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K., Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H., Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs
JOURNAL Unpublished (1999)
COMMENT Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .287
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="2200002K22"
/clone_lib="Mus musculus stomach C57BL/6J adult"
/sex="male"
/tissue_type="stomach"
/dev_stage="adult"
BASE COUNT 73 a 79 c 52 g 82 t 1 others
ORIGIN

Query Match 12.6%; Score 32.6; DB 10; Length 287;

```
  Best Local Similarity   38.1%;  Pred. No. 10;
  Matches   61;  Conservative   19;  Mismatches   80;  Indels    0;  Gaps    0;

Qy       8 cgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaugacguccc 67
           | |  |:   ||  || ||||  : | |  |   |  :| :     |||||:|| |   |
Db     166 CAATAATATGGGTCCAAAACCAGTGCTGTGGTCCCTTTGTTGCCTGGGGGATGAAGGTAC 107

Qy      68 uucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcuguguggggaagc 127
           :  :  | ||:|  || |  | |:|    |:|||:   |  |  ||  |:|    |||
Db     106 TCTTTACACATCTTCATANCCGGTCGGAAGTGGGTTCACATGCAGGGTCGTGGAAAAAGG 47

Qy     128 agucaguauaauucccgucgugugguggugacgccucacga 167
           :   :|: | :||| :   |   | :|   ||: | ||
Db      46 TATGGTTATGACTCCCTTTTGGAAGGCTGGGTCCTGAAGA 7


RESULT   6
AQ657755/c
LOCUS       AQ657755      561 bp    DNA             GSS       23-JUN-1999
DEFINITION  Sheared DNA-21P13.TF Sheared DNA Trypanosoma brucei genomic clone
            Sheared DNA-21P13, DNA sequence.
ACCESSION   AQ657755
VERSION     AQ657755.1  GI:5165523
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 561)
  AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.
            , Fraser,C. and Adams,M.
  TITLE     Determination of clone end sequences from Trypanosoma brucei GUTat
            10.1 sheared DNA library
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: Sheared DNA-21P13.TR
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
            DNA library constructed at TIGR. Clones will be available for
            distribution through ATCC. Sheared DNA end sequences search page:
            http://www.tigr.org/tdb/mdb/tbdb/.
            Seq primer: M13-Forward
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .561
                     /organism="Trypanosoma brucei"
                     /strain="TREU927/4 GUTat 10.1"
                     /db_xref="taxon:5691"
                     /clone="Sheared DNA-21P13"
                     /clone_lib="Sheared DNA"
                     /note="Vector: pUC18; Site_1: SmaI; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from a cloned population of
                     Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
                     sheared to give a tight size distribution (approx 2 kb).
                     The v + i method used for the library construction is
                     described in detail in Smith, H.O. and Venter, J.C.
                     (Making small insert libraries for whole genome shotgun
                     sequencing projects.  In Genome Sequencing: A Practical
                     Approach, eds. M. Vaudin and B. Barell, Oxford University
                     Press, 1999)."
BASE COUNT      164 a    176 c    147 g     74 t
ORIGIN


  Query Match           12.6%;  Score 32.6;  DB 13;  Length 561;
  Best Local Similarity   35.8%;  Pred. No. 12;
  Matches   57;  Conservative   23;  Mismatches   79;  Indels    0;  Gaps    0;

Qy      89 cgucucgggugggugaggaguccuggcuguguggggaagcagucaguauaauuucccgucgu 148
           | :| |    |||   |  |  |: ||: : :| | ||  | | |  :     ||:||:
Db     479 CTTCACCCCAGGGCTCGCCGATCTTGCTCTTTGTGCAGGCGGCTGGGTTGGCTGCGTCGT 420

Qy     149 guguggugacgccucacgacguauuuguccgcugugcagagcguaguaccaagggcugca 208
           |:  ||    | |:  || ||: :  |:|| |: :|        ||:| |       | |
Db     419 GTAGGGGCGTGGCTGTCGGCGTCTCCGTCCTCTTTGTTATTGTCAGTAGCGCCCCGAGTA 360

Qy     209 ccccgguuuuuguuccaagcggagggcaacccccgcuug 247
              |||:::: |: |  |  |    ||  |||  ||::|
Db     359 TGTCGGTTTTAGTGCTTACTGCTCCGCTTCCCGGGCTTG 321


RESULT   7
BB385246/c
LOCUS       BB385246      283 bp    mRNA            EST       13-JUL-2000
DEFINITION  BB385246 RIKEN full-length enriched, 0 day neonate cerebellum Mus
            musculus cDNA clone C230038F17 3' similar to X04663 Mouse mRNA for
            beta-tubulin (isotype Mbeta 5), mRNA sequence.
ACCESSION   BB385246
VERSION     BB385246.1  GI:9107959
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 283)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .283
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="C230038F17"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
```

cerebellum"
/tissue_type="cerebellum"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 479.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 59 a 89 c 56 g 79 t
ORIGIN

Query Match 12.4%; Score 32; DB 10; Length 283;
Best Local Similarity 43.3%; Pred. No. 16;
Matches 58; Conservative 19; Mismatches 55; Indels 2; Gaps 1;

Qy 2 guuuggcgaccaugguggaucagaaccguuucgggugaagccaugucugaaggggauga 61
Db 162 GTTGGAGAAAGCTGAGGGATCTAAGCAGGCTGAGCTGATGCAGGGCTCTGTGGGTTAGGG 103

Qy 62 cgucccuucuggcucauccacaaaaaccgucucgggu--gggugaggaguccuggcugug 119
Db 102 CTCCTCTTCTGGCTCCTCTCGGAAATCCTCTTCGTCTTCCGGGGTGGAATCGTGGTTGGG 43

Qy 120 ugggaagcagucag 133
Db 42 TGGTAGTCGGACAG 29

RESULT 8
AL568674
LOCUS AL568674 1022 bp mRNA EST 16-FEB-2001
DEFINITION AL568674 LTI_FL002_PL1 Homo sapiens cDNA clone CS0DE004YE14 3 prime , mRNA sequence.
ACCESSION AL568674
VERSION AL568674.1 GI:12923250
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1022)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES Location/Qualifiers
source 1. .1022
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DE004YE14"
/clone_lib="LTI_FL002_PL1"
/lab_host="DH10B"
/note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville , Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
BASE COUNT 195 a 283 c 323 g 214 t 7 others
ORIGIN

Query Match 12.2%; Score 31.6; DB 10; Length 1022;
Best Local Similarity 42.8%; Pred. No. 29;
Matches 65; Conservative 12; Mismatches 75; Indels 0; Gaps 0;

Qy 9 gaccaugguggaucagaaccguuucgggugaagccaugucugaaggggaugacgucccu 68
Db 316 GAGCATTGTCCAAGCTGGCTCTTGGGGGTCCCCCATTGGCCACAAGGCCTCACCCCCCA 375

Qy 69 ucuggcucauccacaaaaaccgucucgggugggugaggaguccuggcuguguggggaagca 128
Db 376 CCCCACCCCCGCACCAGAACCACCTTGATTGCAGCAGAAGTGCAGGCAGTGGGTGACGCA 435

Qy 129 gucaguauaauucccgucgugugguggugacgc 160
Db 436 GGGGGCTAAGCTCCCCCAAAGCGGGTKGGGGC 467

RESULT 9
BB500470/c
LOCUS BB500470 321 bp mRNA EST 26-JUL-2000
DEFINITION BB500470 RIKEN full-length enriched, 0 day neonate kidney Mus musculus cDNA clone D630028G21 3', mRNA sequence.
ACCESSION BB500470
VERSION BB500470.1 GI:9489257
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 321)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M. , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y. , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y. , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I. , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino ,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)

Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .321
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D630028G21"
/clone_lib="RIKEN full-length enriched, 0 day neonate kidney"
/tissue_type="kidney"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 89 a 108 c 54 g 70 t
ORIGIN

Query Match 12.1%; Score 31.4; DB 10; Length 321;
Best Local Similarity 46.1%; Pred. No. 25;
Matches 41; Conservative 12; Mismatches 36; Indels 0; Gaps 0;

Qy 55 gggaugacgucccuucuggcucauccacaaaaaccgucucgggugggugaggaguccugg 114
Db 99 GGGATTGTGTGCTTTAAGGGTTACTAAGTAGAATAGGCCGCTGGGGGAGAGGAGTTCTGG 40

Qy 115 cugugugggaagcagucaguauaauuccc 143
Db 39 AGCGGTGGGAAGAAGTGTCCATGATTGCC 11

RESULT 10
BB425694/c
LOCUS BB425694 251 bp mRNA EST 18-JUL-2000
DEFINITION BB425694 RIKEN full-length enriched, 12 days embryo spinal cord Mus musculus cDNA clone C530047N13 3' similar to X04663 Mouse mRNA for beta-tubulin (isotype Mbeta 5), mRNA sequence.
ACCESSION BB425694
VERSION BB425694.1 GI:9265421
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 251)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .251
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="C530047N13"
/clone_lib="RIKEN full-length enriched, 12 days embryo spinal cord"
/tissue_type="spinal cord"
/dev_stage="12 days embryo"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 47 a 72 c 56 g 76 t
ORIGIN

Query Match 12.0%; Score 31.2; DB 10; Length 251;
Best Local Similarity 41.7%; Pred. No. 27;
Matches 45; Conservative 15; Mismatches 48; Indels 0; Gaps 0;

Qy 1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
Db 131 AGTTGGAGAAAGCTGAGGGATCTAAGCAGCCTGAGGTGATGCAGGGGTCTCTGGCTCAGG 72

Qy 61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggag 108
Db 71 CTTCCCCTTCTGGCTCCTCTCCAAAATCGTCTTCCACTGCTAGGGGGG 24

RESULT 11
BB421037/c
LOCUS BB421037 281 bp mRNA EST 16-JUL-2000
DEFINITION BB421037 RIKEN full-length enriched, 12 days embryo spinal cord Mus musculus cDNA clone C530025A17 3' similar to X04663 Mouse mRNA for

```
          beta-tubulin (isotype Mbeta 5), mRNA sequence.
ACCESSION   BB421037
VERSION     BB421037.1  GI:9242392
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 281)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .281
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="C530025A17"
                     /clone_lib="RIKEN full-length enriched, 12 days embryo
                     spinal cord"
                     /tissue_type="spinal cord"
                     /dev_stage="12 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with XhoI and BamHI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT       51 a     82 c     60 g     88 t
ORIGIN


  Query Match             12.0%;  Score 31.2;  DB 10;  Length 281;
  Best Local Similarity   41.7%;  Pred. No. 28;
  Matches   45;  Conservative   15;  Mismatches   48;  Indels    0;  Gaps    0;

Qy       1 aguuuggcgaccaugguggaucagaaccguuucgggugaagccauggucugaaggggaug 60
           ||:: |   |   :|  |||:|  | | |  :  | :|||||   ||:|:       | |
Db     154 AGTTGGAGAAAGCTGAGGGATCTAAGCAGCCTGAGCTGAAGCAGGGGTCTCTGCCTTAGG 95

Qy      61 acgucccuucuggcucauccacaaaaaccgucucgggugggugaggag 108
             |  | |::|:| |:| :|  ||||| |  :  | | : || | || |
Db      94 CCTCCTCTTCTGCCTCCTCTCCAAAATCGCTGCCTGTTCGGGGGGGGG 47


RESULT  12
BF577610/c
LOCUS       BF577610      712 bp    mRNA            EST       12-DEC-2000
DEFINITION  602092476F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4206787 5',
            mRNA sequence.
ACCESSION   BF577610
VERSION     BF577610.1  GI:11651322
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 712)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9768  row: h  column: 20
            High quality sequence stop: 696.
FEATURES             Location/Qualifiers
     source          1. .712
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4206787"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      163 a    204 c    187 g    158 t
ORIGIN


  Query Match             12.0%;  Score 31.2;  DB 11;  Length 712;
  Best Local Similarity   35.3%;  Pred. No. 35;
  Matches   55;  Conservative   23;  Mismatches   78;  Indels    0;  Gaps    0;

Qy      23 agaaccguuucgggugaagccauggucugaaggggaugacgucccuucuggcucauccac 82
           ||||    ::   |:  |   |: |:     |    :| |    |::|: |  | :||||
Db     413 AGAATGTCTTGCCGTACAAAAATTGTGCATGGTCAGTGGCACATCTTCTAGGACCTCCAC 354

Qy      83 aaaaaccgucucgggugggugaggaguccuggcuguguggggaagcagucaguauaauucc 142
           ||  | |  |:|       |||| |   :|||:|:|:   | ||||: | :    ::
Db     353 AATGAGCTCCTCACCATCCAGAGGCGCAATGGCTGTGTCCCAGGCAGTGACTTGTCTTGT 294

Qy     143 cgucgugugguggugacgccucacgacguauuugucc 178
           | :| : :|  | | |     || ||  |   |:||
```

```
Db    293 CTTCCTTTGATGAGCCTGTAGACCACACAGCAGTCC 258


RESULT  13
BE647687
LOCUS       BE647687      275 bp    mRNA            EST       06-SEP-2000
DEFINITION  UI-M-BH1-anm-c-11-0-UI.r1 NIH_BMAP_M_S2 Mus musculus cDNA clone
            UI-M-BH1-anm-c-11-0-UI 5', mRNA sequence.
ACCESSION   BE647687
VERSION     BE647687.1  GI:9973500
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 275)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
            should be noted that Bento Soares is generating a small number of
            additional specialized non-redundant arrays of BMAP cDNAs whose
            availability will be considered under appropriate and limited
            collaborative arrangements
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .275
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH1-anm-c-11-0-UI"
                     /clone_lib="NIH_BMAP_M_S2"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S2 library is a subtracted library derived from
                     NIH_BMAP_M_S1, which in turn is a subtracted library
                     derived from a mixture of normalized libraries from ten
                     regions of the mouse brain (cerebellum, brain stems,
                     olfactory bulbs, hypothalamus, cortex, amygdala, basal
                     ganglia, pineal gland, striatum, hipoccampus). The driver
                     used for subtraction consisted of a pool of 5,000 clones
                     from the NIH_BMAP_M_S1 library and a pool of 2,000 clones
                     obtained from non-normalized and normalized mouse brain
                     spinal cord libraries."
BASE COUNT       61 a     53 c     95 g     66 t
ORIGIN


  Query Match             12.0%;  Score 31;  DB 10;  Length 275;
  Best Local Similarity   38.4%;  Pred. No. 32;
  Matches   78;  Conservative   25;  Mismatches   95;  Indels    5;  Gaps    1;

Qy      35 ggugaagccauggucugaaggggaugacgucccuucuggcucauccacaaaaaccgucuc 94
           || | |  ||: |  :|    |  : ||    |::| ||| |       |  |  |   |
Db      63 GGGGTATTCATAGGGTGTGTAGATTAACCCAGCTTCAGGCCCTGGAGGGACTATTGCAGC 122

Qy      95 gggugggugaggaguccuggcuguguggggaagcagucaguauaauuccccguc-----gug 149
             | :|||:||||||: | |  :|    |  ||| |:| |:|:   :|   :|
Db     123 AGTTGGGTGAGGAGTTCCGTTTGGCATGACAGCGGTCTGTATTTGTCTGATCAAAGGCAT 182

Qy     150 uguggugacgccucacgacguauuuguccgcugugcagagcguaguaccaagggcugcac 209
```

```
           : :||:   |||    | ||:|  :|: ||| | ||||   |: | |  |  |||
Db     183 TATGGTAGGGCCAGCTGGCGTAAGTGTACGCAGAGCAGCTGGTGGGAGGACAGGCGCAGG 242

Qy     210 cccgguuuuuguuccaagcggag 232
           ||| |:  :  ::|   | | ||
Db     243 CCCAGTGATGATTCTTGGAGCAG 265


RESULT  14
AW361210
LOCUS       AW361210      548 bp    mRNA            EST       04-FEB-2000
DEFINITION  RC1-CT0252-231099-013-h09 CT0252 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW361210
VERSION     AW361210.1  GI:6865860
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 548)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC1&t2=RC1-CT0252-
            231099-013-h09&t3=1999-10-23&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 85
            High quality sequence stop: 547.
FEATURES             Location/Qualifiers
     source          1. .548
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CT0252"
                     /dev_stage="Adult"
                     /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      124 a    118 c    150 g    156 t
ORIGIN


  Query Match             12.0%;  Score 31;  DB 10;  Length 548;
  Best Local Similarity   39.4%;  Pred. No. 38;
  Matches   80;  Conservative   23;  Mismatches   95;  Indels    5;  Gaps    1;

Qy      35 ggugaagccauggucugaaggggaugacgucccuucuggcucauccacaaaaaccgucuc 94
           || | |  ||: |  :| | |  | : |      |::| ||| |       ||  |  |   |
Db     127 GGGGTACTCATAGGGTGTATAGATTAAACCAGCTTCGGGCCCTGGAGGAACTATTGCAGC 186

Qy      95 gggugggugaggaguccuggcuguguggggaagcagucaguauaauuccccguc-----gug 149
             | :|||:||||||: | |  :|    |  ||| |:| |:|:   :|   :|
Db     187 AGTTGGGTGAGGAGTTCCGTTTGGCATGACAGCGGTCTGTATTTGTCTGATCAAAGGCAT 246

Qy     150 uguggugacgccucacgacguauuuguccgcugugcagagcguaguaccaagggcugcac 209
           : :||:   |||    | ||:|   |: ||| | ||||   |: | |  || |||
Db     247 TATGGTAGGGCCAGCTGGCGTAGGAGTACGCAGGGCAGCTGGTGGGAGAAACGGCGCAGG 306

Qy     210 cccgguuuuuguuccaagcggag 232
           ||| |:  :  : |   | | ||
```

```
Db    307 CCCAGTAATGATCCTTGGAGCAG 329


RESULT  15
BH110789
LOCUS       BH110789      586 bp    DNA             GSS       19-JUL-2001
DEFINITION  RPCI-24-352N22.TV RPCI-24 Mus musculus genomic clone RPCI-24-352N22
            , DNA sequence.
ACCESSION   BH110789
VERSION     BH110789.1  GI:14945264
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 586)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-352N22.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 352  row: N  column: 22
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .586
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-352N22"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI partially digested male C57BL/6J
                     DNA."
BASE COUNT      160 a    131 c    110 g    185 t
ORIGIN


  Query Match             12.0%;  Score 31;  DB 13;  Length 586;
  Best Local Similarity   39.1%;  Pred. No. 38;
  Matches   34;  Conservative   18;  Mismatches   35;  Indels    0;  Gaps    0;

Qy    145 ucgugugguggugacgccucacgacguauuuguccgcugugcagagcguaguaccaagggc 204
          : |:|:| | : | |  :| | |||  : :|:|| |:|   || || : |:| || ||
Db    262 TTGTGTGGGCTCAGGGTTCTCTACGGCTATGTCCACTGCAGAGTGCATTGTAGCATGGCG 321

Qy    205 ugcaccccgguuuuuguuccaagcgga 231
          :   |    |::: : ::| | || ||
Db    322 TTAGCATTCGTTTGTCTTCTAGGCAGA 348


Search completed: February 16, 2002, 15:47:47
Job time: 6410 sec
```